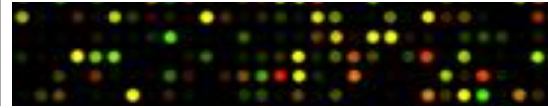




UNIVERSIDAD AUTÓNOMA
DE MADRID

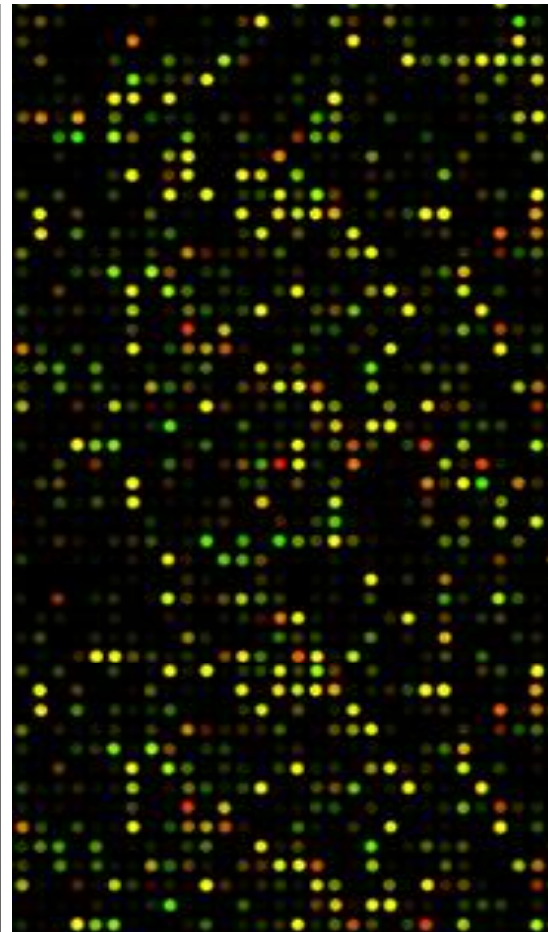
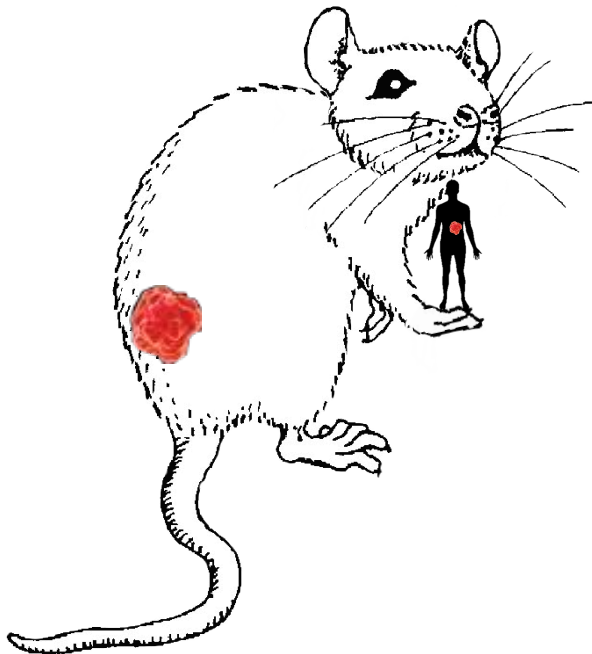
Departamento de Biología Molecular

Facultad de Ciencias



Análisis transcripcional de xenoinjertos derivados de pacientes con cáncer de páncreas

TESIS DOCTORAL



Raquel Martínez García

Madrid, 2014



UNIVERSIDAD AUTÓNOMA DE MADRID

FACULTAD DE CIENCIAS

Departamento de Biología Molecular



**ANÁLISIS TRANSCRIPCIONAL DE XENOINJERTOS DERIVADOS DE
PACIENTES CON CÁNCER DE PÁNCREAS**

Tesis Doctoral que presenta la licenciada en Ciencias Biológicas

Raquel Martínez García

realizada en el Centro Nacional de Investigaciones Oncológicas

Directores:

Dr. Alfonso Valencia Herrera
Dr. Manuel Hidalgo Medina
Dr. Daniel Rico Rodríguez



D. Alfonso Valencia Herrera, Director del Programa de Biología Estructural y Biocomputación, **D. Manuel Hidalgo Medina**, Director del Programa de Investigación Clínica y **D. Daniel Rico Rodríguez**, Investigador del Grupo de Biología Computacional Estructural,

CERTIFICAN: Que Dña. Raquel Martínez García ha realizado el trabajo original de investigación “Análisis transcripcional de xenoinjertos derivados de pacientes con cáncer de páncreas”, bajo su dirección en el Centro Nacional de Investigaciones Oncológicas, para la obtención del Grado de Doctor.

Que consideran que dicho trabajo reúne las condiciones necesarias para su presentación y defensa ante tribunal en la Facultad de Ciencias de la Universidad Autónoma de Madrid.

Y para que así conste, a efectos oportunos, firman el presente certificado.

Madrid, a 1 de Septiembre de 2014

Alfonso Valencia Herrera

Manuel Hidalgo Medina

Daniel Rico Rodríguez

“Los ordenadores son inútiles. Sólo pueden darte respuestas.”

Picasso

AGRADECIMIENTOS

La presente tesis es un esfuerzo en el cual, directa o indirectamente, han participado muchas personas opinando, dando ánimo, acompañando en los momentos de crisis y en los momentos de felicidad.

Agradezco a los Dres. Manuel Hidalgo y Alfonso Valencia haber confiado en mí y permitirme realizar la tesis bajo su dirección, así como por sus consejos, el apoyo y el ánimo que me han brindado. También agradezco al Dr. Daniel Rico haber asumido la co-dirección de este trabajo y, por supuesto, su paciencia y su incalculable colaboración en este trabajo. A David de Juan por su ayuda y comentarios en todo el proceso de elaboración de la tesis. Al Dr. Pedro P. López por las largas charlas y discusiones sobre todos los temas que se me hayan podido ocurrir, manteniendo siempre una paciencia infinita.

Gracias también a mis queridos compañeros, que me apoyaron y me permitieron entrar en su vida durante estos casi cinco años de convivir dentro y fuera del CNIO. Paolo, Ángel, Kristina, Marta, Jorge, Camino y Natalia, gracias.

A mis padres y a mi hermana que me han acompañado en esta aventura de forma incondicional. A ti Oriol, que desde un principio hasta el día de hoy sigues dándome ánimo para terminar este proceso.

Gracias a todos.

RESUMEN

El injerto de muestras de adenocarcinoma ductal de páncreas (ADP) en ratones para generar xenoinjertos derivados de paciente (PDX) es una plataforma muy prometedora para estudios biológicos y terapéuticos de esta enfermedad. Sin embargo, estos modelos todavía no están completamente caracterizados. En este trabajo medimos el impacto del ambiente murino en la expresión génica de las células humanas tumorales injertadas.

Primero, desarrollamos un protocolo de análisis robusto para explorar los espacios de expresión génica usando perfiles de modelos PDX, tumores primarios y líneas celulares de ADP y carcinoma hepatocelular (CHC). Además, estudiamos la contribución de las células no tumorales a los perfiles de expresión y los espacios. Hemos completado el análisis con una caracterización funcional de los modelos PDX, incluyendo si los cambios son causados por el ambiente murino o por los diferentes pases. Finalmente, comparamos los modelos PDX derivados de tumor primario y de metástasis.

Nuestro trabajo tiene tres implicaciones importantes. Primero, los modelos PDX de ADP (y CHC) mantienen similitudes a nivel de expresión con los tumores primarios, mientras que las líneas celulares derivadas de ADP y CHC son indistinguibles. Segundo, los cambios de expresión en los modelos PDX están relacionados principalmente con rutas biológicas que reflejan la pérdida de células humanas infiltrantes y la adaptación a un nuevo ambiente. Además, demostramos la estabilidad de los patrones de expresión génica a lo largo de los pases, indicando que los cambios se producen en las fases tempranas del proceso de adaptación. Tercero, los modelos PDX derivados de tumor primario y metástasis tienen perfiles de expresión muy similares y muestran una respuesta comparable frente a un conjunto de fármacos utilizados en cáncer de páncreas.

ABSTRACT

Engraftment of primary pancreas ductal adenocarcinomas (PDAC) in mice to generate patient-derived xenograft (PDX) models is a promising platform for biological and therapeutic studies in this disease. However, these models are still incompletely characterized. In this study we measured the impact of the murine tumor environment on the gene expression of the engrafted human tumoral cells.

First, we developed a robust analysis protocol to explore the gene expression space using profiles of PDX models, primary tumors and cell lines from PDAC and hepatocellular carcinoma (HCC). In addition, we study the contribution of non-tumoral cells to the expression profiles and the spaces. We completed the analysis with a functional characterization of PDX models, including if changes were caused by murine environment or by serial passing. Finally, we compared the PDX models derived from primary tumors and from metastasis.

Our work has three important implications. First, PDAC (and HCC) PDX models retain gene expression similarities with the primary tumors, while cell lines derived from PDAC and HCC are indistinguishable. Second, expression changes in PDXs are mainly related to pathways reflecting the lack of human infiltrating cells and the adaptation to a new environment. We also provide evidence of the stability of gene expression patterns over subsequent passages, indicating early phases of the adaptation process. Third, PDX models derived from primary tumor or metastasis show very similar gene expression profiles and comparable drug response sensitivity against a set of pancreatic cancer drugs.

ÍNDICE

Abreviaturas.....	i
1. Introducción	1
1.1. Adenocarcinoma ductal de páncreas (ADP)	1
1.1.1. Anatomía del páncreas y tipos de cáncer asociados	1
1.1.2. Lesiones precursoras del ADP y sus alteraciones genéticas	2
1.1.3. Alteraciones genéticas y epigenéticas del ADP	4
1.1.4. Estadios de evolución del ADP	7
1.1.5. Tratamiento del ADP	10
1.2. Invasión y metástasis en otros tejidos	10
1.3. Los xenoinjertos derivados de paciente (PDX) como modelos en estudios preclínicos e investigación	13
1.3.1. Estatus genómico de los modelos PDX	16
1.4. Perfiles de expresión génica	18
1.5. Rutas biológicas y cáncer	18
2. Hipótesis y Objetivos	21
3. Materiales y métodos	23
3.1. Protocolo de generación de los modelos PDX	23
3.2. Muestras usadas en la tesis	24
3.2.1. Datos generados en el estudio	24
3.2.2. Datos preexistentes	24
3.3. Protocolo de análisis de perfiles de expresión	25

3.3.1. Corrección del fondo y normalización.....	26
3.3.2. Determinación de ausencia o presencia de transcritos (<i>expression calls</i>) ..	27
3.3.3 Generación de espacios de expresión mediante MCA	27
3.3.4 Análisis complementarios al MCA.....	28
3.4. Cuantificación del componente murino en los perfiles de expresión de modelos PDX	29
3.5. Cuantificación del componente no tumoral	30
3.6. Estimación y corrección del componente no tumoral	30
3.7. Análisis de enriquecimiento funcional.....	31
3.7.1. Análisis basado en frecuencias de expresión	31
3.7.2. Análisis basado en t-test sobre las intensidades.....	32
3.8. Expresión diferencial entre diferentes pases	33
3.9. Estudio de eficacia de un fármaco	33
3.10. Predicción de los reguladores transcripcionales asociados a los modelos PDX.....	34
4. Resultados.....	37
4.1. Modelos PDX derivados de tumor primario	37
4.1.1. Espacio de expresión génica de ADP	37
4.1.2. Espacio de expresión génica de carcinoma hepatocelular (CHC)	41
4.1.3. Espacios de expresión génica combinando muestras de ADP y CHC.....	43
4.1.4. Contribución de las células no tumorales infiltrantes al perfil de expresión de los modelos PDX	50
4.1.5. Efecto de la microdissección en la contribución de las células no tumorales infiltrantes a los perfiles de expresión de los tumores primarios.....	60
4.1.6. Análisis funcional de los modelos PDX de ADP.....	69
4.1.7. Estabilidad de los perfiles de expresión de los modelos PDX.....	72

4.2. Modelos PDX derivados de metástasis	77
4.2.1. Análisis de los grupos funcionales en metástasis humanas	77
4.2.2. Espacios de expresión génica de modelos PDX derivados de tumor y de metástasis	78
4.2.3. Respuesta a fármacos de los modelos PDX derivados de tumor primario y de metástasis	84
4.2.4. Predicción de los reguladores transcripcionales asociados a modelos PDX	86
5. Discusión	91
5.1. Modelos PDX de ADP derivados de tumor primario	91
5.1.1. El modelo PDX como modelo de investigación y para estudios preclínicos	91
5.1.2. Los modelos PDX mantienen una memoria de expresión génica de los tumores de origen.....	92
5.1.3. Las células no tumorales infiltrantes contribuyen a la subestimación de las similitudes entre los modelos PDX y los tumores primarios	93
5.1.4. Las alteraciones funcionales se producen por la adaptación al microambiente murino y se mantienen estables a lo largo de los pases	95
5.2. Modelos PDX de ADP derivados de metástasis	97
5.2.1. Los modelos PDX y las metástasis comparten alteraciones funcionales	97
5.2.2. Los modelos PDX derivados de tumor primario son indistinguibles de los modelos PDX derivados de metástasis.....	98
5.2.3. Los modelos PDX derivados de tumor primario y los modelos PDX derivados de metástasis parecen mostrar un sensibilidad a fármacos similar	98
5.2.4. Factores de transcripción con actividad diferencial en modelos PDX	100

6. Conclusiones	105
Bibliografía	107
Anexo I – Material suplementario	125
Anexo II – Publicaciones y comunicaciones a congresos.....	139

ABREVIATURAS

ACVR1B: *activin A receptor, type IB*

ADP: adenocarcinoma ductal de páncreas

ARID1A: *AT rich interactive domain 1A*

ARID2: *AT rich interactive domain 2*

ATM: *ataxia telangiectasia mutated*

BRAF: *v-raf murine sarcoma viral oncogene homolog B*

BRCA1: *breast cancer 1*

BRCA2: *breast cancer 2*

CDKN2A/p16: *cyclin-dependent kinase inhibitor 2A*

CHC: carcinoma hepatocelular

CLDX: xenoinjertos derivados de líneas celulares (*cell lines-derived xenograft*)

CNA: alteración en el número de copia (*copy number alteration*)

CNIO: Centro Nacional de Investigaciones Oncológicas

CSC: célula madre del cáncer (*cancer stem cell*)

CTC: célula tumoral circulante (*circulating tumoral cell*)

CTNNB1: *catenin (cadherin-associated protein), beta 1*

EHF: *ets homologous factor*

EPC1: *enhancer of polycomb homolog 1*

ESTIMATE: *estimation of stromal and immune cells in malignant tumours using expression data*

FDR: tasa de falso descubrimiento (*false discovery rate*)

Fn: pase n (siendo n el número de pase del modelo PDX, 0 es el tumor primario)

fRMA: *frozen robust multiarray analysis*

FT: factor de transcripción

GEO: *gene expression omnibus (database)*

GHM: Grupo Hospital de Madrid

GO: *gene ontology*

GSEA: análisis de enriquecimiento de conjuntos de genes (*gene set enrichment analysis*)

ICH: índice Calinsky y Harabsz

IPMN: neoplasias mucinosas intraductales (*intraductal papillary mucinous neoplasm*)

ISMARA: *integrated system for motif activity response analysis*

JAK1: *janus kinase 1*
 JH: *John Hopkins Sidney Kimmel Cancer Center*
 KEGG: *Kyoto encyclopedia of genes and genomes*
 KRAS: *kristen rat sarcoma viral oncogen homolog*
 MARA: *motif activity response analysis*
 MCA: análisis de correspondencias múltiples (*multiple correspondence analysis*)
 MCN: neoplasias mucinosas quísticas (*mucinous cystic neoplasm*)
 miRNAs: *microRNAs*
 MLL3: *histone-lysine N-methyltransferase*
 MYF: *myogenic factor*
 NES: valor normalizado de enriquecimiento (*normalized enrichment score*)
 NFATC: *nuclear factor of activated T-cells, cytoplasmic*
 NFY: *nuclear transcription factor Y*
 NTI: (células) no tumorales infiltrantes
 PALB2: *partner and localizer of BRCA2*
 PanIn: neoplasias intraepiteliales pancreáticas (*pancreatic intraepithelial neoplasia*)
 PCA: análisis de componentes principales (*principal component analysis*)
 PDX: xenoinjerto derivado de paciente (*patient-derived xenograft*)
 PTEN: *phosphatase and tensin homolog*
 RMA: *robust multiarray analysis*
 SF3B1: *splicing factor 3b, subunit 1, 155kDa*
 SMAD4/DPC4: *SMAD family member 4*
 SNAI: *snail family zinc finger*
 SPF: libre de gérmenes patógenos (*specific pathogen free*)
 STK11: *serine/threonine kinase 11*
 STR: repeticiones cortas en tándem (*short tandem repeat*)
 TGF- β : *transforming growth factor, beta*
 TGFBR1: *transforming growth factor, beta receptor 1*
 TGFBR2: *transforming growth factor, beta receptor 2*
 TFBS: sitio de unión a factor de transcripción (*transcription factor binding site*)
 TGI: tasa de inhibición del crecimiento tumoral (*tumor growth inhibition*)
 TP53: *tumor protein p53*
 UI: unidad internacional
 WNT: *wingless-type MMTV integration site family*

1. Introducción

1.1. Adenocarcinoma ductal de páncreas (ADP)

1. 1. 1. Anatomía del páncreas y tipos de cáncer asociados

El páncreas es un órgano retroperitoneal con funciones de glándula exocrina y endocrina. Como se observa en la Figura 1, el páncreas consta de diferentes partes: la cabeza (íntimamente relacionada con el duodeno y el conducto biliar común) y el cuerpo y la cola (situadas en posición ascendente hacia el bazo). La función exocrina está desempeñada por células epiteliales dispuestas en estructuras esféricas u ovoides huecas llamados acinos pancreáticos. La sección exocrina es la encargada de la secreción de enzimas digestivas. La parte endocrina se agrupa en Islotes de Langerhans, que consisten en cúmulos de células secretoras de hormonas que producen insulina, glucagón y somatostatina.

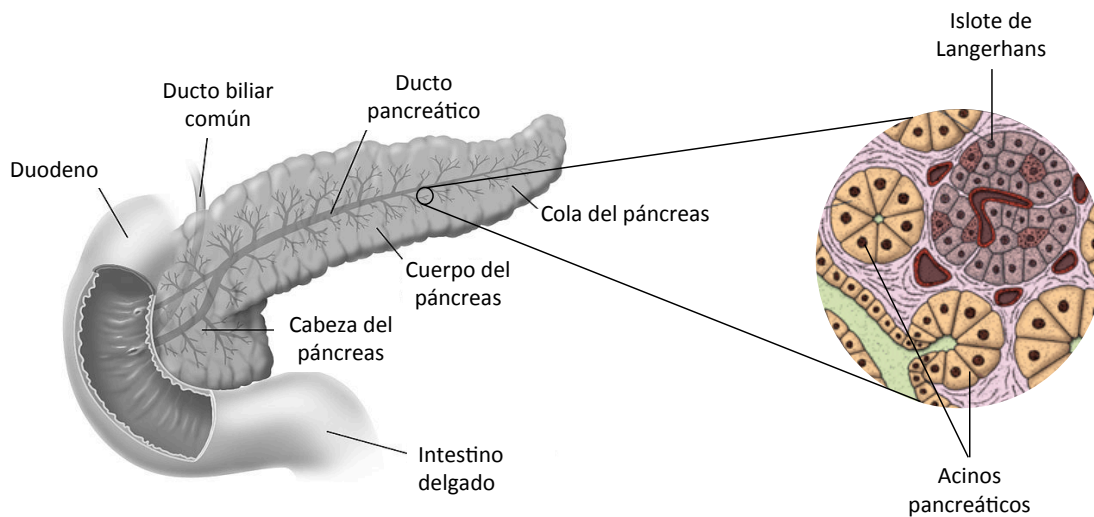


Figura 1. Esquema anatómico del páncreas y detalle ampliado, donde se observan las principales células exocrinas y endocrinas que lo componen. Adaptado de <http://www.yourcancertoday.com/Cancers/Pancreatic-Cancer/44> y <http://simplegreen1591.blogspot.com.es/2013/04/sistema-digestivo.html>.

El cáncer de páncreas es la cuarta causa de muerte por cáncer en países desarrollados [38]. En España se diagnostican más de 5000 nuevos casos cada año (fuente: GLOBOCAN 2008, *International Agency for Research on Cancer*). Existen diversos tipos histopatológicos de tumores malignos en el páncreas, siendo el adenocarcinoma ductal de páncreas (ADP) la variante que representa el tipo histológico más frecuente (80-90%). Por ello, es habitual que el término “cáncer de páncreas”, generalmente, se refiera al ADP. Así pues, en esta tesis, siempre que nos referimos al cáncer de páncreas estaremos hablando solo de esta forma mayoritaria. Los otros tipos de cáncer de páncreas menos frecuentes se muestran en la Tabla 1.

Tipo histopatológico	Origen	Frecuencia
Adenocarcinoma ductal de páncreas (ADP)	Exocrino	80-90%
Carcinoma de células acinares	Exocrino	2%
Cistoadenocarcinoma mucinoso	Exocrino	3%
Cistoadenocarcinoma seroso	Exocrino	<1%
Neoplasia papilar mucinosa intraductal	Exocrino	1-3%
Carcinoma pseudopapilar	Exocrino	<1%
Pancreatoblastoma	Exocrino	Raro
Gastrinoma	Endocrino	1-2%
Insulinoma	Endocrino	1-2%
Glucagonoma	Endocrino	<1%
Somatostatinaoma	Endocrino	Raro
VIPoma	Endocrino	<1%

Tabla 1. Clasificación histopatológica de los diferentes tipos de cáncer de páncreas [21, 31, 61].

1.1.2. Lesiones precursoras del ADP y sus alteraciones genéticas

En el año 2000, Hruban et. al propusieron un modelo de progresión de ADP, según el cual las células de los ductos evolucionan transformándose en diferentes lesiones precursoras [41]. Dichas lesiones son provocadas por la acumulación de una serie de

cambios genéticos en la células ductales del páncreas [41] (Figura 2). El primer cambio genético consiste en la mutación del oncogén KRAS. Asimismo, el acortamiento de los telómeros también se ha descrito como uno de los eventos tempranos implicados en la tumorigénesis del ADP. A continuación, en lo que ya se puede considerar lesiones de alto grado, se produce la inactivación del gen supresor de tumores CDKN2A/p16. Finalmente, se producen los eventos más tardíos, que incluyen la pérdida de genes supresores de tumores (TP53 y SMAD4/DPC4) y genes de mantenimiento del genoma (BRCA2) [41, 73].

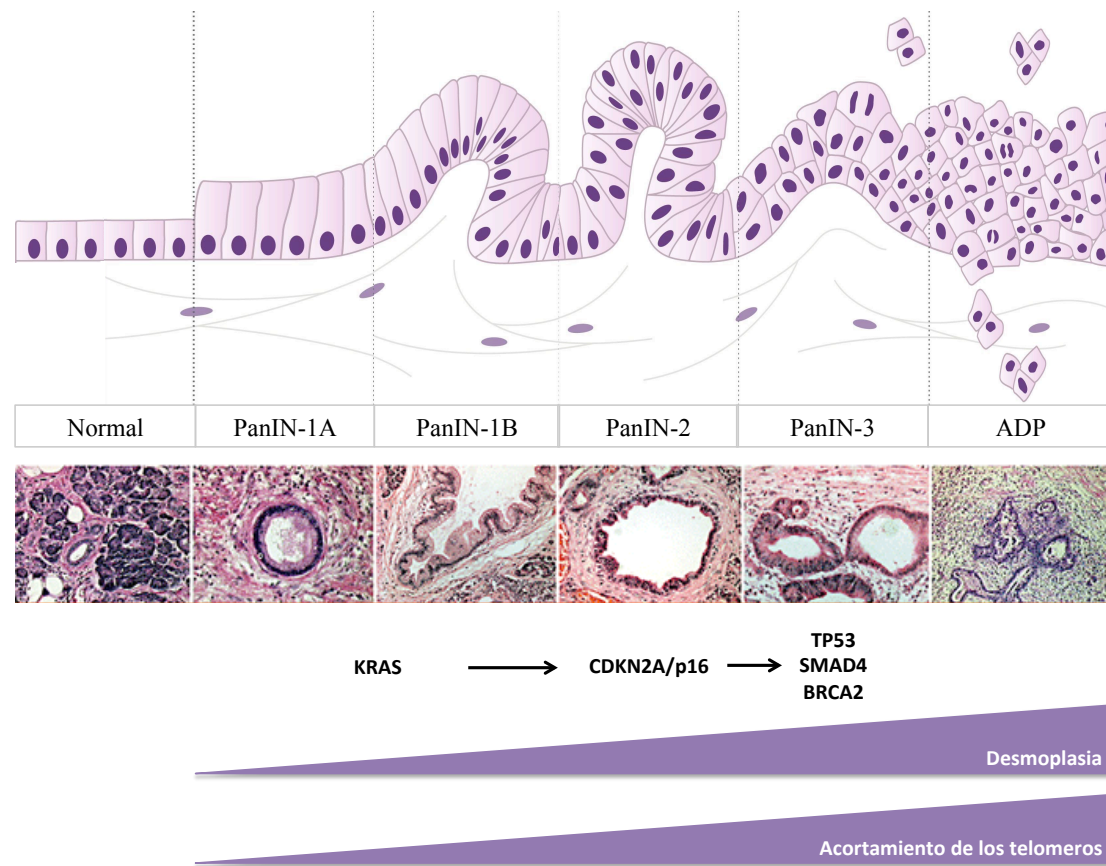


Figura 2. Modelo de progresión del ADP basado en la evolución de las neoplasias intraepiteliales pancreáticas (PanIN) hasta dar lugar al ADP. En la parte intermedia de la figura se puede observar un muestra histológica de cada paso correspondiente de la evolución. En la parte inferior se muestran las mutaciones adquiridas a lo largo de dicha evolución. Adaptado de Chang et al. (2008) [13].

Las posibles lesiones precursoras que pueden derivar en ADP son: las neoplasias mucinosas intraductales (IPMN, *Intraductal Papillary Mucinous Neoplasm*), las neoplasias mucinosas quísticas (MCN, *Mucinous Cystic Neoplasm*) y las neoplasias intraepiteliales pancreáticas (PanIN, *Pancreatic Intraepithelial Neoplasia*) [42]. Las IPMN son neoplasias no invasivas productoras de mucina que crecen formando prolongaciones filiformes. Dichas prolongaciones pueden ocupar el conducto pancreático principal o sus ramificaciones. Se localizan con mayor frecuencia en la cabeza del páncreas. Por su parte, las MCN son neoplasias que pueden asociarse a un componente invasivo. Son productoras de mucina como las IPMN, sin embargo, se localizan preferentemente en el cuerpo y cola del páncreas. Por último, las PanIN son lesiones no invasivas, localizadas principalmente en los ductos pancreáticos de la cabeza y la cola. Hay tres tipos de lesiones según su grado de atipia (PanIN-1, PanIN-2 y PanIN-3). En la Figura 2 se puede observar un esquema de la evolución de estas lesiones, así como unas muestras histológicas de las mismas. Adicionalmente, en la Tabla 2 se pueden consultar la frecuencia de mutación, en cada una de las lesiones precursoras, de los genes pertenecientes al modelo de progresión.

Gen	IPMN	MCN	PanIN	Referencias
KRAS	50-100%	90%	80-100%	[50, 58, 126, 128]
CDKN2A/p16	80%		50-90%	[126]
TP53	25-50%	50%	70%	[50, 85, 102, 126]
SMAD4/DPC4	10-100%	100%	0-30%	[47, 48, 126]
BRCA2	4%		0-10%	[68, 126]

Tabla 2. Frecuencia de mutación de los genes presentes en el modelo de progresión de Hruban et al. (2000) en las diferentes lesiones precursoras.

1.1.3. Alteraciones genéticas y epigenéticas del ADP

La genómica está teniendo un importante auge en los últimos años, gracias a las tecnologías de secuenciación cada vez más sofisticadas. Además, los avances en bioinformática permiten realizar análisis de genomas completos.

Una de las principales alteraciones genómicas estudiadas en ADP son las mutaciones puntuales. Las mutaciones más frecuentes en el ADP son aquellas que se han descrito previamente en el modelo de progresión, que incluye el oncogén KRAS (90%) y los genes supresores de tumores CDKN2A/p16 (95%), TP53 (75%) y SMAD4/DPC4 (55%) [94]. Los valores de las frecuencias de mutación varían según la fuente consultada, pero en todas ellas KRAS es el gen más frecuentemente mutado. No obstante, también se conocen genes mutados en una menor proporción de los casos (5%-10%) y entre los cuales hay genes supresores de tumores (BRCA2, PTEN, STK11, TGFBR1, TGFBR2, ACVR1B), implicados en los procesos de *splicing* (SF3B1), proteínas quinasas relacionadas con cascadas de señalización (JAK1, BRAF), modificadores/remodeladores de cromatina (ARID1A, ARID2, EPC1, MLL3) y genes relacionados con los mecanismos de reparación del ADN (ATM) [7, 49, 73]. Muchos de los genes mutados en ADP, solo lo están en un porcentaje bajo de los pacientes, aun así, esas mutaciones pueden ser determinantes para encontrar un tratamiento efectivo y personalizado. En este sentido, es interesante mencionar el primer caso en el que la secuenciación del exoma de un tumor sirvió como base a una decisión clínica específica. En este caso, se detectó una mutación en el gen PALB2 que era potencialmente capaz de inhibir la reparación del ADN, vía su interacción con BRCA1 y BRCA2. Esta información fue utilizada para crear un modelo PDX del tumor, en el contexto de un proyecto de medicina personalizada. Sobre el modelo murino se probaron diferentes tratamientos, incluyendo algunos indicados para mutaciones en BRCA1 y BRCA2. Finalmente, en el modelo PDX resultó efectivo el tratamiento con Mitomicina C. El paciente respondió de forma positiva (con remisión completa) al mismo fármaco. Este caso puede servir como experiencia piloto para ilustrar como la utilización de modelos PDX en el proceso de selección de fármacos (basándose en información genómica), puede ser clave en el tratamiento de tumores como el pancreático [119]. En el contexto de esta tesis, este caso inicial sirve también para ilustrar la importancia del estudio comparativo de la genómica y respuesta a fármacos, de los tumores y modelos PDX derivados de los mismos.

Junto a las mutaciones puntuales, otra de las aberraciones más relevantes en pacientes de ADP son las alteraciones en el número de copia (CNA, *Copy Number Alteration*).

La acumulación de este tipo de alteraciones puede desembocar en la inactivación de genes supresores de tumores y la amplificación de oncogenes, contribuyendo a la iniciación y progresión del cáncer [35, 59]. En la mayoría de los casos de cáncer de páncreas, el genoma muestra grandes regiones duplicadas o perdidas, implicando en algunos casos hasta el 40% de un cromosoma. Las aberraciones en el número de copia son de potencial importancia clínica. En este sentido, existen asociaciones significativas entre la pérdida de copias en los cromosomas 9p y 18q y la metastatización en regiones linfáticas, o entre las deleciones en el cromosoma 4q y pronóstico negativo de ADP [69].

También existen otros factores relacionados con el control de la expresión génica, y en particular con la metilación del ADN, que juegan un papel crítico en el desarrollo y la diferenciación celular [54]. Los patrones de metilación del ADN están implicados en diversas enfermedades, incluido el cáncer [26]. La hipermetilación de las islas CpG está implicada en el silenciamiento de genes supresores de tumores [37], mientras que su hipometilación se ha relacionado con una aumento en la expresión de oncogenes [19]. En el estudio de Nones et al. (2014) [87], además, se muestra que el 33% de los genes diferencialmente metilados, están inversamente correlacionados con los niveles de expresión de los genes. Esos genes están implicados en rutas biológicas previamente asociadas a ADP, como por ejemplo, TGF- β , WNT o adhesión celular.

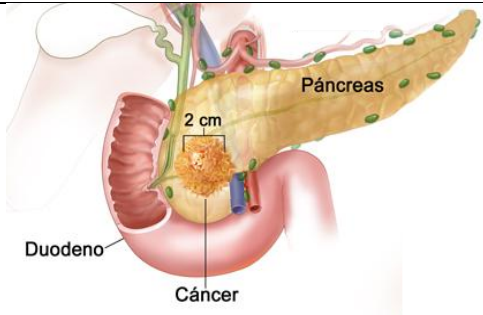
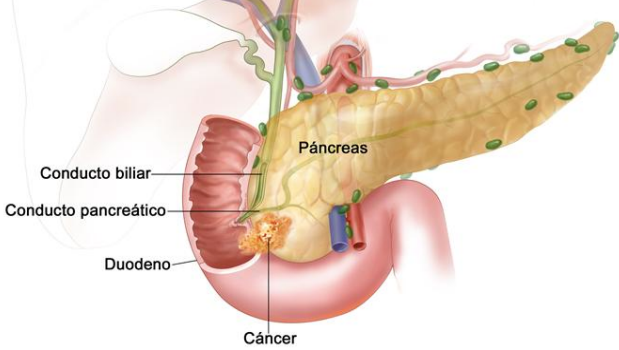
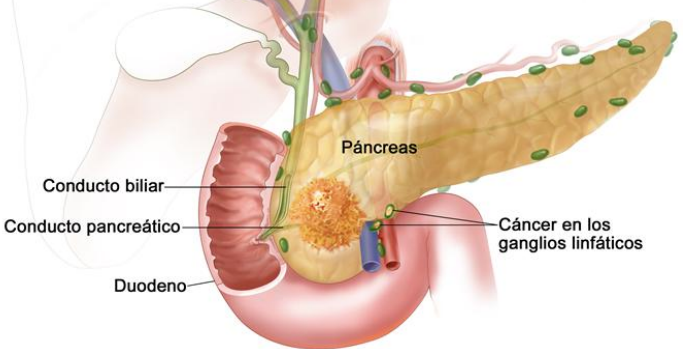
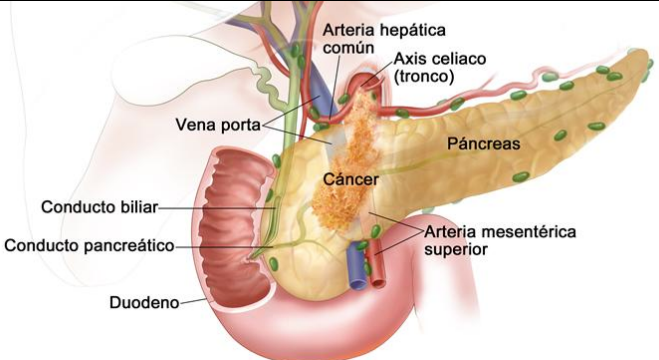
Finalmente, en los últimos años también se han realizado varios esfuerzos para analizar el papel en cáncer de los *microRNAs* (miRNAs). Los miRNAs están potencialmente implicados en la regulación de la expresión de, aproximadamente, el 30% de la proteínas codificantes. Además, sus distintas combinaciones están relacionadas con proliferación, supervivencia, resistencia a fármacos y metástasis [107]. A su vez, la propia expresión de los miRNAs se regula mediante modificaciones epigenéticas, alteraciones en el número de copia o mutaciones genéticas [65]. Los miRNAs también se han utilizado en ADP como biomarcador de diagnóstico. Por ejemplo, se estudió que el uso combinado de miR-16 con el marcador CA19-9, previamente utilizado en ADP, produce un aumento significativo de la sensibilidad y especificidad en el diagnóstico de cáncer de páncreas [29].

1.1.4. Estadios de evolución del ADP

Frente al diagnóstico de ADP es importante definir la capacidad de resecabilidad del tumor, es decir, la posibilidad de realizar cirugía sobre el tumor y la probabilidad de éxito asociada. Para realizar la clasificación se utilizan los criterios TNM, establecidos por el grupo *International Union Against Cancer* [105]. Los criterios TNM son los siguientes:

- T: se refiere al tumor primario. Se puede subdividir a su vez en:
 - T1: el tumor se limita al páncreas y mide como máximo 2 cm en su dimensión mayor.
 - T2: el tumor se limita al páncreas y mide más de 2 cm en su dimensión mayor.
 - T3: el tumor se extiende más allá del páncreas, sin compromiso del tronco celíaco o la arteria mesentérica superior.
 - T4: el tumor compromete el tronco celíaco o la arteria mesentérica superior.
 - TX: no se puede evaluar.
- N: hace referencia a la presencia de invasión tumoral en ganglios linfáticos regionales.
 - N0: no hay invasión tumoral en los ganglios linfáticos regionales.
 - N1: si hay invasión tumoral en los ganglios linfáticos regionales.
 - NX: no se pueden evaluar los ganglios linfáticos regionales.
- M: se refiere a la presencia de metástasis a distancia.
 - M0: no hay metástasis a distancia.
 - M1: si hay metástasis a distancia.

Los tumores se clasifican en estadios según los criterios TNM. En la Tabla 3 se muestran los estadios que puede presentar el ADP, su correspondiente código TNM y una ilustración de su localización.

Estadio	TNM	Ilustración
IA IB	T1N0M0 T2N0M0	
IIA	T3N0M0	
IIB	T1N1M0 T2N1M0 T3N1M0	
III	T4, cualquier N, M0	

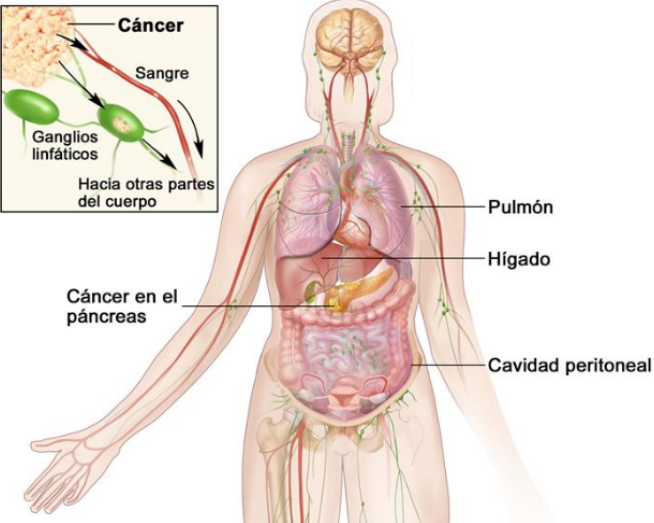
IV	Cualquier T, cualquier N, M1	
----	------------------------------------	--

Tabla 3. Clasificación de los estadios del ADP según los criterios TNM. Adaptado de <http://www.cancer.gov/espanol/pdq/tratamiento/pancreas/HealthProfessional/page3>.

Según la clasificación TNM, un tumor pancreático se considera:

- Resecable: tumor con crecimiento local que permite la cirugía con resección completa del tumor. Incluye los estadios T1N0M0, T2N0M0 y T3N0M0 (si la afectación vascular es mínima). Únicamente el 20% de los pacientes diagnosticados son candidatos a tratamiento mediante cirugía (esto hace especialmente difícil la obtención de material biológico para investigación del ADP).
- Localmente avanzado: el tumor ha crecido afectando e invadiendo los vasos y ganglios linfáticos regionales. La cirugía no permite la resección completa del tumor respecto de los vasos o ganglios. Incluye los estadios T1-3N1M0 y T4N0-1M0. El 30% de los pacientes presentan este tipo de características al momento del diagnóstico.
- Metastático: el tumor ha migrado y colonizado otro órgano. La cirugía del tumor primario no tiene valor sin la cirugía de las metástasis. Incluye los estadios T1-4N0-1M1. En el momento del diagnóstico, aproximadamente en el 50% de los pacientes se ha producido la invasión y metastatización, principalmente en hígado o pulmón [110].

Para este trabajo los tumores resecables se consideran y se denominan tumores primarios, mientras que los tumores localmente avanzados y metastáticos se engloban como tumores metastáticos.

1.1.5. Tratamiento del ADP

Actualmente el tratamiento más eficaz es la cirugía, debido al bajo porcentaje de éxito que la mayoría de los tratamientos farmacológicos presentan. En este aspecto, se están investigando diversas estrategias basadas en la utilización de quimioterapia, con o sin radioterapia, administrada de forma adyuvante o neoadyuvante. Los tratamientos de referencia en la actualidad son el 5-Fluororacilo, la Gemcitabina y el Nab-paclitaxel (quimioterápicos aplicados tras la cirugía).

La supervivencia de los pacientes que han sido tratados con cirugía es de aproximadamente 2 años, cantidad que desciende dramáticamente hasta los 8 meses en los pacientes en los cuales no se puede llevar a cabo cirugía [34, 39]. Por ello, es totalmente necesario obtener terapias alternativas.

1.2. Invasión y metástasis en otros tejidos

La propagación de un tumor puede llevarse a cabo mediante dos mecanismos diferentes: invasión o metástasis [86, 89]. La invasión es la migración y penetración directa de las células del cáncer en los tejidos circundantes. Sin embargo, la metástasis implica la capacidad de las células tumorales de penetrar en los vasos sanguíneos o linfáticos, desplazarse a través de la circulación sanguínea y después crecer en un nuevo tejido anatómicamente distante. Por ello, los tumores que se extienden a los ganglios linfáticos cercanos al tumor primario (ganglios linfáticos regionales) se consideran tumores localmente avanzados.

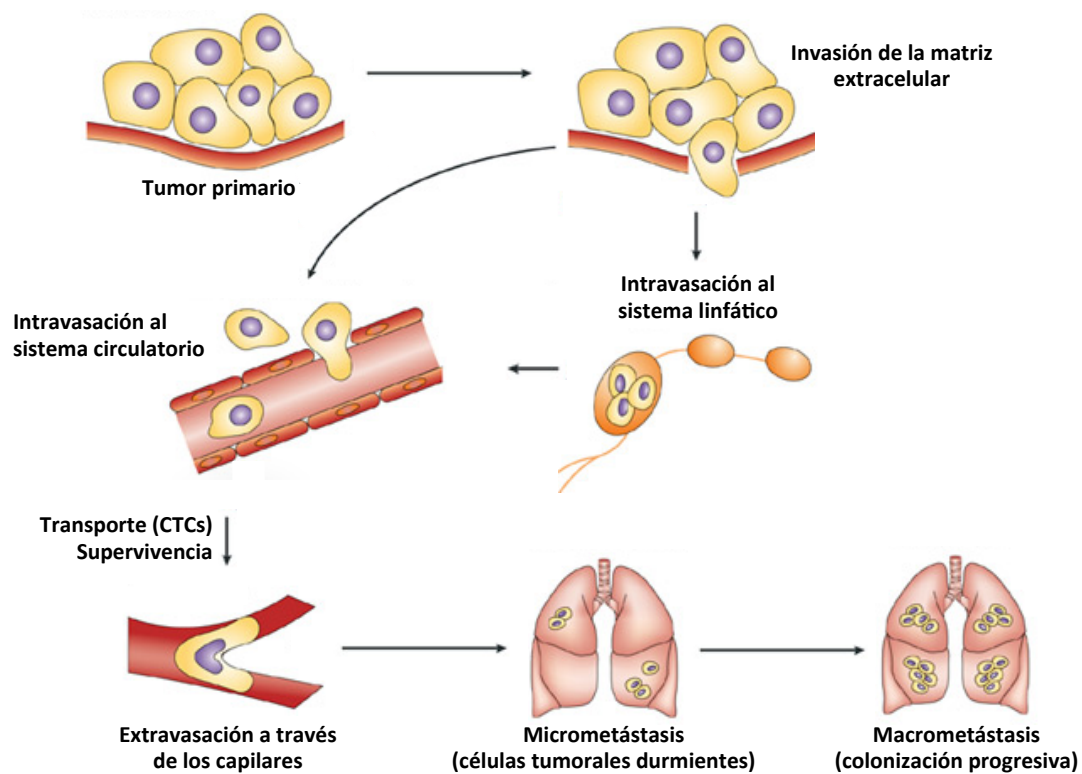


Figura 3. Esquema de la cascada metastática [108] mediante la cual un tumor primario deriva en una metástasis.

La metástasis se produce mediante una serie de pasos, biológicamente complejos, denominados cascada metastática (Figura 3) [115, 118]. En dicha cascada, las células tumorales abandonan el lugar original del tumor y emigran a otras partes del cuerpo, a través de la circulación sanguínea o linfática. La cascada metastática se inicia mediante un proceso de degradación e invasión de la matriz extracelular. A continuación se produce la intravasación, fenómeno por el cual la célula tumoral se introduce en un vaso sanguíneo o linfático y procede a su circulación por el organismo. Cuando la célula se encuentra en el torrente sanguíneo se denomina célula tumoral circulante (CTC). Las células normales que se encuentran en la circulación poseen un tamaño y una elasticidad características, que les permite mantenerse en movimiento hasta llegar a su lugar de destino. Debido a que las CTCs no poseen estas características, las restricciones de tamaño y elasticidad provocan la detención de las CTCs en los capilares [115, 118].

Otro mecanismo que puede ocasionar la detención de las CTCs es el tropismo metastático. Las localizaciones más frecuentes de las metástasis son los órganos más irrigados por la sangre, como son el cerebro, los pulmones, el hígado, los huesos y las glándulas suprarrenales. Los riñones y el corazón constituyen la excepción a esta regla, dado que no presentan una mayor incidencia de metástasis, a pesar de ser órganos altamente irrigados [86]. Sin embargo, también es conocida la tendencia de ciertos tumores a diseminarse en determinados órganos, este fenómeno se conoce como tropismo metastático. Este tropismo puede deberse a la expresión de moléculas de adhesión en las células tumorales cuyos ligandos se encuentran en las células del órgano diana, o a la liberación de sustancias quimiotácticas en el órgano diana que recluten a las células tumorales [57]. El cáncer de páncreas, aunque puede diseminarse en cualquier órgano, tiende a propagarse en hígado y pulmón [110]. El siguiente paso de la cascada metastática es la extravasación de la célula tumoral a la matriz circundante del nuevo órgano. Finalmente, las células tumorales pueden permanecer durmientes como micrometástasis hasta que comienza la colonización progresiva, que conlleva la proliferación en el nuevo tejido para formar una macrometástasis. Para que se produzca dicha proliferación es necesario el desarrollo de nuevos vasos sanguíneos, proceso denominado angiogénesis [115, 118].

Otra teoría bastante aceptada entre la comunidad científica, se basa en la idea de que la generación de metástasis se produce a partir de células madre cancerosas (CSCs, *Cancer Stem Cell*). Existen diferentes hipótesis sobre el origen de las CSCs, pero en todos los casos se postula que una o varias mutaciones activan la reprogramación celular. Sin embargo, la controversia reside en qué tipo celular constituye el origen de las CSCs. Algunos autores proponen que derivan de células madre multipotentes [2], otros autores de células progenitoras tejido específicas (también denominadas células madre unipotentes) [66] y, por último, hay autores que proponen que derivan de células maduras diferenciadas [8, 74]. Independientemente del origen, las CSCs se consideran parte de la población del tumor. Estas células podrían provocar la reincidencia de un tumor primario o llevar a cabo los pasos de la cascada metastática y dar lugar a metástasis (Figura 4).

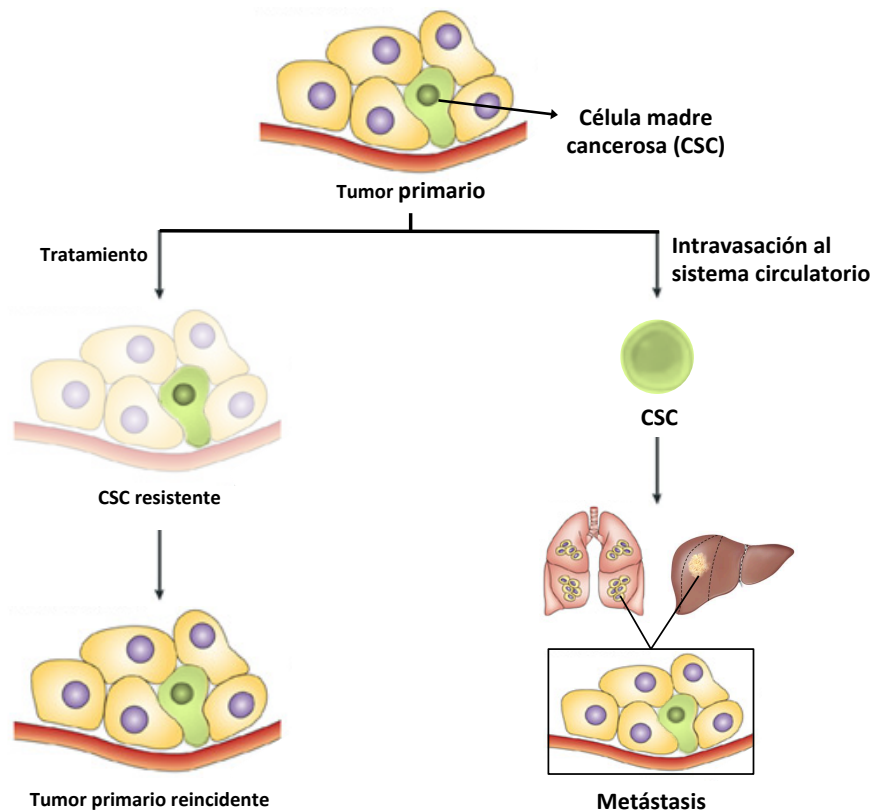


Figura 4. Esquema de la generación de metástasis a partir de CSCs. Adaptado de Jordan et al. (2006) [56].

1.3. Los xenoinjertos derivados de paciente (PDX) como modelos en estudios preclínicos e investigación

Los estudios preclínicos son pruebas que se llevan a cabo en un contexto científicamente controlado, con utilización de cultivos celulares y animales como modelos. La finalidad de los estudios preclínicos, es analizar la interacción y mutuas actuaciones entre el organismo y el candidato a fármaco. En los estudios de tumores humanos, las herramientas más utilizadas para llevar a cabo un estudio preclínico son las siguientes:

1. Líneas celulares. Son células tumorales cultivadas *in vitro* en condiciones controladas.

2. Xenoinjertos derivados de líneas celulares (CLDXs, *Cell Lines-Derived Xenografts*). Se generan inyectando células tumorales humanas, previamente cultivadas *in vitro*, en modelos animales (principalmente ratón).
3. Xenoinjertos derivados de pacientes (PDXs, *Patient-Derived Xenografts*). Se implanta directamente una porción del tumor del paciente en el modelo animal, principalmente ratón.

Las líneas celulares y los modelos animales permiten realizar experimentos de respuesta a fármacos de un modo rápido y efectivo, por lo que son sistemas imprescindibles en los estudios preclínicos. Además, estos modelos permiten eliminar, o al menos limitar, la contaminación de células humanas no neoplásicas que, en distinta proporción, siempre están presentes en los tumores humanos. La principal desventaja de estos modelos, es el hecho de que no pueden reproducir las condiciones asociadas al tejido de soporte que rodea al tumor (estroma), compuesto por fibroblastos, células del sistema inmunitario y elementos vasculares. El estroma tiene un papel crítico en la proliferación celular y la metástasis [46, 95]. A favor de los modelos animales puede decirse que ambos modelos de xenoinjertos, mantienen una arquitectura tisular consistente [22, 101, 113] gracias a la sustitución de las células estromales humanas por las murinas [14, 62, 106].

El principal problema de los CLDXs es que el cultivo celular, previo al injerto, provoca una homogeneidad genética celular que puede no reflejar aspectos importantes de la heterogeneidad del tumor primario. La heterogeneidad genética se refiere a la diversidad existente entre tumores de diferentes individuos (diversos subtipos moleculares) o en el mismo paciente (diferentes regiones o poblaciones celulares de un mismo tumor) [83]. La homogeneidad genética en las líneas celulares, es debida a que el proceso de cultivo celular lleva asociado una selección de las células que mejor se adaptan a las condiciones *in vitro* [121]. En general, los resultados obtenidos usando modelos PDX muestran mejor concordancia con resultados clínicos y preclínicos que los resultados obtenidos utilizando CLDXs [53, 79, 98]. Además, los estudios con CLDXs muestran un bajo poder de predicción de respuesta a fármacos [52, 62, 71].

Por todas las razones expuestas, los modelos PDX se están convirtiendo en una plataforma de uso habitual con propósitos clínicos y de investigación [40]. Estos modelos se han incluido en diversos estudios sobre la biología y farmacología de cáncer de mama [22], cáncer renal [103] o glioblastoma [24]. El injerto de tumores humanos es, actualmente, el único método disponible que nos permite la propagación de manera significativa de un carcinoma [18, 45, 99]. Asimismo, los modelos PDX constituyen una plataforma esencial para la creación de un biobanco, es decir, un banco *in vivo* de tumores [97]. A pesar de la elevada letalidad que presenta el ADP, un biobanco permite la utilización de muestras tumorales incluso tras el fallecimiento de los pacientes.

Los modelos PDX pueden ser implantados de forma subcutánea u ortotópica. En los modelos subcutáneos el tumor se implanta bajo la piel de ratón, mientras que en los ortotópicos el tumor se implanta en el mismo lugar anatómico en el cual se originó. Por tanto, los modelos PDX ortotópicos pueden mantener gran parte de las interacciones fisiológicas con el órgano en el que se desarrollan. Sin embargo, tienen un mayor coste debido a las dificultades técnicas quirúrgicas necesarias para la implantación, y a que son necesarias técnicas de imagen para la evaluación y medición del tumor.

En nuestro laboratorio, se utilizan modelos PDX implantados de forma subcutánea, de forma habitual como herramienta de estudio. Por ello, en adelante utilizaremos el término modelo PDX para referirnos a modelos de xenoinjertos derivados de paciente e implantados de manera subcutánea. Los modelos PDX se generan a medida que los pacientes llegan a los hospitales colaboradores y dan su consentimiento. En este trabajo, se incluyen los modelos PDX disponibles durante el desarrollo de la tesis para los que se generaron perfiles de expresión.

Los modelos PDX se generan a partir de un fragmento de tumor del paciente. Tras la extracción del tumor del paciente (F0), la fase de implantación (F1) es la primera fase. Para poder realizar diversos estudios se llevan a cabo fases de expansión (F2 - Fn), que permiten obtener el número de ratones necesarios para cada estudio (eficacia de

un tratamiento, búsqueda de nuevos biomarcadores...). Cada vez que se reimplanta un tumor se considera un nuevo pase ($F_n + 1$). En la Figura 5 se puede observar un esquema de las diferentes fases.

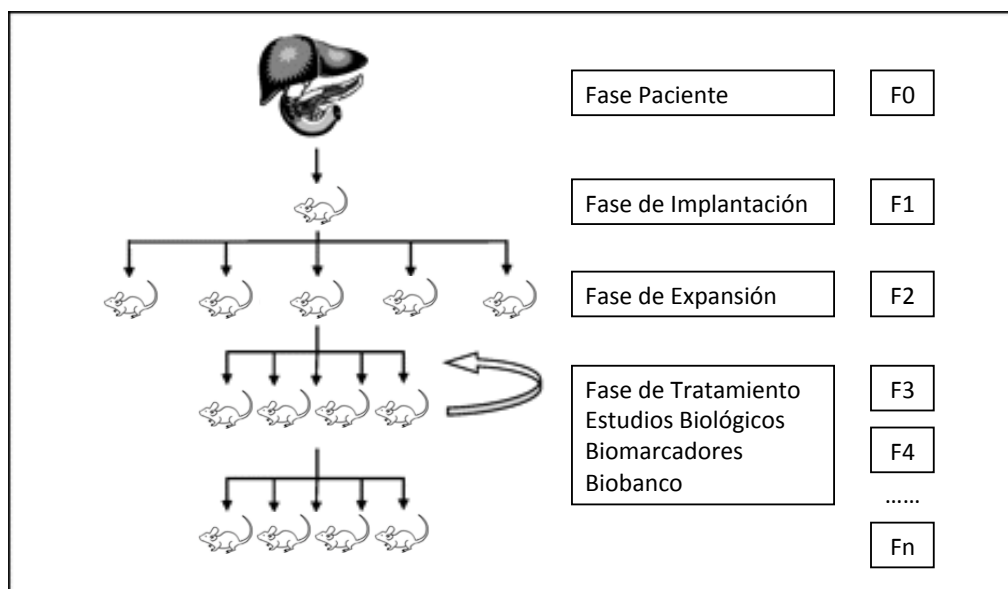


Figura 5. Esquema de fases del proceso de generación de un modelo PDX [98].

1.3.1. Estatus genómico de los modelos PDX

Entre los genes comúnmente mutados en ADP (KRAS, CDKN2A/p16, TP53, SMAD4/DPC4 y BRCA2, ver sección 1.1.3.), KRAS y SMAD4 han sido seleccionados por varios grupos para comprobar la estabilidad genómica de los modelos PDX.

En el estudio de Rubio-Viqueira et al. (2006) [98] se usaron modelos PDX de ADP. Las muestras analizadas pertenecían a pase 0 (F0), es decir, tumor primario procedente del paciente, y pase 3 (F3) del modelo PDX generado a partir del mismo paciente. La Tabla 4 muestra que la mayoría de los modelos en F3 mantenían el estado mutacional de KRAS y la expresión de SMAD4, respecto al tumor humano original.

Paciente	KRAS		SMAD4/DPC4	
	F0	F3	F0	F3
Panc163	GAT/wt	GAT/wt	+	+
Panc185	GAT/wt	GAT/wt	-	-
Panc194	GTT/wt	GTT/wt	-	-
Panc198	GAT/wt	GAT/wt	-	-
Panc215	GAT/wt	GAT/wt	+	+
Panc219	GAT/wt	GAT/wt	+	+
Panc253	GAT/wt	GAT/wt	+	+
Panc265	GAT/wt	GAT/wt	Heterogéneo	+
Panc281	GTT/wt	GTT/GTT	-	-
Panc286	GAT/wt	GAT/wt	+	+
Panc287	wt/wt	wt/wt	-	-
Panc294	GTT/wt	GTT/GTT	+	-

Tabla 4. Comparación de la estabilidad de los modelos PDX mediante el estudio del estatus de KRAS y la expresión de SMAD4. Código: wt, *wild type*; +, expresado; -, no expresado. Adaptado de Rubio-Viqueira et al. (2006) [98].

En un trabajo reciente, se ha realizado un estudio con 8 tumores primarios de ADP y sus correspondientes modelos PDX [79]. En ese trabajo se llevó a cabo una caracterización molecular de los modelos PDX a muchos niveles, incluyendo análisis de número de copia, expresión génica, miRNA, mutaciones, STR (*Short Tandem Repeat*) e inmunohistoquímica. Los modelos PDX pertenecían a diferentes países, incluyendo algunos modelos con más de tres países. En el análisis de número de copia, se encontró una estabilidad a lo largo de los países, debido a que las alteraciones recurrentes estaban presentes en países tempranos y tardíos. Asimismo, los perfiles de expresión de los tumores primarios y de los modelos PDX mostraban una alta correlación. La caracterización de las mutaciones se realizó seleccionando los genes KRAS, CTNNB1, TP53 y SMAD4. En la mayoría de los casos, las muestras procedentes del mismo paciente mostraban las mismas alteraciones, incluso en los modelos PDX con un número de países muy elevado [79].

Los modelos PDX han sido usados en otros tipos de cáncer, como son el cáncer de cuello de útero [14], cáncer microcítico de pulmón [20], cáncer de vejiga [88], cáncer de mama [95] o melanoma ocular uveal [64]. En el caso de Fichtner et al. (2008) [27] se generaron modelos PDX implantando tumores de pulmón no microcítico. Este estudio reveló un alto grado de similitud, en los perfiles de expresión, entre los tumores primarios y sus correspondientes modelos PDX.

1.4. Perfiles de expresión génica

Los perfiles de expresión génica son una de las herramientas más útiles para hacer comparaciones de muestras diferentes pero relacionadas, como pueden ser el tumor primario procedente del paciente y su correspondiente modelo PDX. Los perfiles de expresión, habitualmente se generan mediante hibridación de ARN en *microarrays* de ADN o por secuenciación masiva de ARN (*RNAseq*). El uso de forma simultánea de un elevado número de perfiles de expresión, procedentes de diferentes muestras, permite realizar un meta-análisis transcripcional.

En este sentido, en un estudio de Gadaleta et al. (2011) [28] se analizó lo que denominaron espacio pancreático de expresión (*pancreas expression space*), utilizando tumores primarios de cáncer de páncreas, muestras de tejido no tumoral adyacente al tumor, tejido pancreático de un donante sano, líneas celulares y modelos PDX. Los resultados de este estudio se discuten en el contexto de nuestros propios resultados en esta tesis.

1.5. Rutas biológicas y cáncer

La investigación del cáncer durante mucho tiempo, se ha basado en la detección de los genes implicados en la tumorigénesis, pero en la actualidad también existen numerosos esfuerzos centrados en el entendimiento del metabolismo del cáncer y de las rutas biológicas implicadas [33]. El tumor establece relaciones, tanto a nivel

interno como con las células del tejido adyacente. A partir de esas interacciones, se desencadenan señalizaciones para la secreción o inhibición de genes y proteínas, desembocando en diferentes procesos biológicos que favorecen la situación tumoral. Por ejemplo, cuando un tumor crece y se empieza a generar necrosis, la célula empieza a detectar una disminución de la cantidad de oxígeno disponible [43]. En ese momento, la célula tumoral activa la ruta de producción de lactato (efecto Warburg [121]), vía activación del gen Ras. El gen Ras activado impulsa la angiogénesis y favorece la invasividad tumoral.

Existen diversas bases de datos centradas en recopilar toda la información disponible acerca de las rutas biológicas. Entre las más populares se encuentran:

- KEGG (*Kyoto Encyclopedia of Genes and Genomes*): base de datos de rutas metabólicas. Consta de rutas pertenecientes a cientos de especies procariotas y eucariotas.
<http://www.genome.jp/kegg/pathway.html>
- REACTOME: base de datos que recopila la gran mayoría de procesos biológicos celulares humanos, incluyendo información metabólica.
<http://www.reactome.org>
- Gene Ontology (GO): base de datos de características biológicas (proceso biológico, localización celular y función molecular) en las que está implicado cada gen del genoma de diversas especies.
<http://www.geneontology.org>

El estudio de las rutas implicadas en la progresión del tumor, es tan importante como el estudio de los genes. Por ejemplo, en el trabajo de Jones et al. (2008) [55], se demostró que las muestras de tumor de páncreas analizadas contenían una media de 63 alteraciones genéticas, siendo la mayoría mutaciones puntuales. La mayor parte de las mutaciones presentaban una frecuencia muy baja, es decir, eran alteraciones individuales en cada paciente. Solo los genes KRAS, CDKN2A/p16 y TP53 se encontraban mutados en la mayoría de los individuos del estudio. Curiosamente, el análisis funcional demostró que los genes mutados estaban implicados,

fundamentalmente, en 12 rutas biológicas. Las mutaciones eran individuales, pero las rutas alteradas eran comunes en la mayoría de los tumores analizados (67-100%). Esto significa que cada paciente puede tener diferentes mutaciones en un gen o en diversos genes, pero esas mutaciones pueden afectar a determinadas rutas biológicas, importantes en la generación y progresión del tumor.

Este tipo de situaciones también ocurren en otras enfermedades, como por ejemplo la leucemia linfática/linfocítica crónica. Quesada et al. (2013) [92], llevaron a cabo un análisis de mutaciones a partir de los datos de secuenciación de diversas muestras de este tipo de leucemia. Dicho análisis mostró que la mayoría de las mutaciones detectadas poseían una baja frecuencia y solo unos pocos genes estaban mutados en más del 10% de los pacientes. El análisis funcional, sin embargo, reveló que los genes mutados estaban implicados, principalmente, en 6 rutas biológicas [77, 91].

Por ello, en esta tesis llevaremos a cabo diversos estudios de enriquecimiento funcional, facilitando la interpretación de los perfiles de expresión de las células cancerígenas en su ambiente original y en los modelos PDX.

2. Hipótesis y Objetivos

El principal objetivo de la presente tesis doctoral es realizar un análisis exhaustivo de los cambios que se producen a nivel transcripcional en las células humanas procedentes de adenocarcinomas ductales de páncreas (ADP), cuando son injertadas en un ratón para establecer un modelo PDX (*Patient-Derived Xenograft*). Para ello, hemos definido los siguientes objetivos específicos:

1. Generar perfiles de expresión de modelos PDX y procesarlos junto con datos de expresión obtenidos de bases de datos públicas.
2. Desarrollar un protocolo robusto para generar espacios de expresión génica que permita el análisis sistemático y cuantitativo de los perfiles de expresión.
3. Caracterizar las funciones moleculares y las redes de control transcripcional que diferencian los modelos PDX de las correspondientes muestras humanas.
4. Estudiar la estabilidad de los patrones de expresión en los modelos PDX.
5. Comparar los patrones de expresión de los modelos PDX derivados de tumor primario y derivados de metástasis.
6. Comparar la respuesta a un conjunto de fármacos de los modelos PDX derivados de tumor primario y derivados de metástasis.

3. Materiales y métodos

3.1. Protocolo de generación de los modelos PDX

Los modelos PDX se generan a partir de un fragmento de tumor del paciente. La muestra de tumor se recoge en el servicio de anatomía patológica del hospital, tras la cirugía o biopsia del paciente. Para su traslado, se introduce en un medio compuesto de RPMI y antibióticos (penicilina, 0.2 UI/ml (siendo UI, Unidad Internacional), y estreptomicina, 0.2 µg/ml) y se mantiene a 4°C. En el menor tiempo posible, el tumor se transporta al animalario del Centro Nacional de Investigaciones Oncológicas (CNIO) para realizar el injerto. Los modelos PDX han sido establecidos siguiendo un protocolo de implantación, previamente descrito y publicado [98], que contiene los siguientes pasos:

1. Fragmentar el tumor humano en trozos de 2-3 mm³ e introducir en Matrigel.
2. Realizar dos pequeñas incisiones en la piel del ratón (*nude* (nu/nu, Foxn1), atímico, de 5 a 6 semanas de edad). Cada incisión debe estar situada en la parte baja de la espalda del animal y se realiza una en cada flanco. Crear un pequeño bolsillo subcutáneo.
3. Implantar un pequeño fragmento en cada bolsillo y cerrar.

Para comprobar el éxito de la implantación se realizan mediciones de forma periódica con un pie de rey. Se considera que el tumor está creciendo cuando alcanza los 100 mm³. Para expandir la colonia y obtener el siguiente pase, se realiza una reimplantación cuando el volumen supera los 1500 mm³. El modelo PDX se considera establecido cuando hemos sido capaces de reimplantarlo al menos tres veces.

Los ratones se mantienen en las instalaciones de animalario en condiciones de barrera SPF (libre de gérmenes patógenos, por sus siglas en inglés, *Specific Pathogen Free*). Además, los ratones usados en este trabajo han sido tratados de forma humanitaria siguiendo la regulación determinada por el Comité Bioético del CNIO y las directivas legales establecidas (86/609/EEC).

3.2. Muestras usadas en la tesis

3.2.1. Datos generados en el estudio

Para esta tesis se han generado 35 nuevos perfiles de expresión de modelos PDX y 2 de tumores primarios. Las muestras de tumor primario proceden de pacientes de los hospitales del Grupo Hospital de Madrid (GHM). Sin embargo, en el caso de los modelos PDX, las muestras y perfiles de expresión proceden de diversos lugares: 11 muestras y sus correspondientes perfiles de expresión proceden de *John Hopkins Sidney Kimmel Cancer Center* (JH), 18 muestras proceden de JH pero sus perfiles de expresión fueron generados específicamente para este estudio y, finalmente, hay 6 pacientes de GHM cuyos modelos PDX fueron establecidos para el biobanco de CNIO y cuyos perfiles fueron generados para esta tesis. Todos estos datos han sido publicados [76] y puestos a disposición de la comunidad científica a través de la base de datos *Gene Expression Omnibus Datasets* (GEO Datasets [6, 23]. Perteneciente al *National Center for Biotechnology Information*. <http://www.ncbi.nlm.nih.gov/geo/>), y se puede acceder a ellos mediante el código GSE51798. Adicionalmente, hemos generado perfiles de expresión de 8 metástasis humanas y 9 modelos PDX derivados de metástasis, todos ellos derivados de pacientes del GHM.

El procesado para la obtención de los perfiles de expresión de todas estas muestras ha sido el mismo. Inicialmente, la extracción del ARN se realiza mediante la herramienta comercial *Qiagen RNeasy Mini kits* (QIAGEN Inc., Valencia, California, USA). A continuación, se procede a la hibridación en la plataforma *GeneChip® Human Genome U133 Plus 2.0 Array* (Affymetrix, Santa Clara, California, USA).

3.2.2. Datos preexistentes

Además de los nuevos perfiles de expresión, específicamente creados para este estudio, hemos utilizado datos de GEO Datasets cuyas muestras han sido hibridadas en la misma plataforma que se hemos usado para nuestros modelos (*GeneChip®*

Human Genome UI33 Plus 2.0 Array). Se han recopilado los siguientes perfiles de expresión:

- Tumores primarios humanos de ADP (GSE15471 [3] y GSE16515 [90])
- Tumores primarios humanos microdisecccionados de ADP (GSE17891 [16])
- Líneas celulares de ADP (GSE21654 [80] y GSE17891 [16])
- Modelos PDX de ADP (GSE9599 [51])
- Modelos PDX de ADP en pase 5 y 10 y el correspondiente tumor primario [30]
- Modelos PDX ortotópicos de ADP (GSE46385 [120])
- Tumores primarios humanos de carcinoma hepatocelular (CHC) (GSE6222 [67], GSE6465 [44], GSE9843 [15])
- Líneas celulares de CHC (GSE6465 [44] y GSE36133 [5])
- Modelos PDX de CHC (GSE6465 [44])
- Metástasis de ADP y tumores primarios (GSE42952 [117] y GSE34153 [12]). Las muestras de GSE34153 han sido hibridadas en la plataforma *Agilent 4x44K Whole Human Genome Array*, y debido a ello, se utilizaron los datos disponibles en la base de datos (datos previamente normalizados por los autores).

Para facilitar la lectura de los distintos resultados se ha incluido la Tabla S1 (Anexo I – Material suplementario), que indica los datos utilizados en cada uno de los análisis presentados.

3.3. Protocolo de análisis de perfiles de expresión

En esta tesis hemos generado nuevos modelos PDX y sus correspondientes perfiles de expresión y, además, hemos recopilado un elevado número de datos públicos de diversas muestras. Para poder analizar todo este volumen de datos y hacerlo de manera robusta, hemos desarrollado un protocolo de análisis seleccionando el método más adecuado en cada paso (Figura 6).

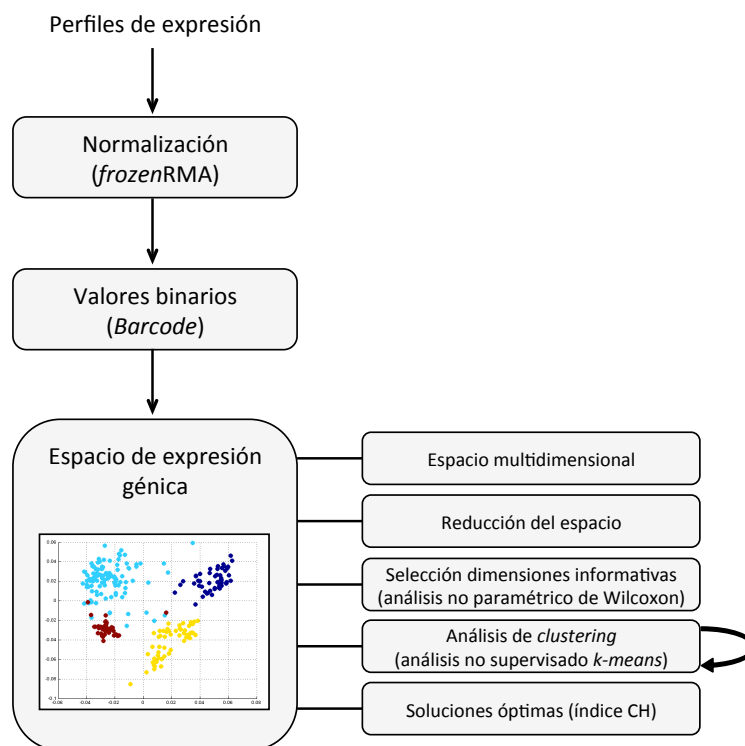


Figura 6. Esquema resumen del protocolo de análisis establecido para el procesado de perfiles de expresión y estudio de los espacios de expresión.

3.3.1. Corrección del fondo y normalización

Primero se lleva a cabo el procesado y normalización de los datos de expresión. Uno de los métodos más usados para el procesado de datos obtenidos con la plataforma Affymetrix, es el RMA (de sus siglas en inglés, *Robust Multiarray Analysis*). RMA se encarga de hacer una corrección del fondo (*background correction*) y una normalización, pero requiere que el conjunto total de datos sea analizado simultáneamente. Como ya se ha mencionado, nuestro laboratorio va generando los modelos PDX de forma continua para su inclusión en el biobanco y para diferentes experimentos. En esta tesis solo se han generado los perfiles de expresión de parte de los modelos PDX disponibles, y por ello, hemos decidido usar un método alternativo llamado *frozenRMA* (fRMA, versión 1.12.0) [81]. Este método nos permite procesar las muestras usando RMA, pero analizando los datos de expresión de cada muestra individualmente o en pequeños grupos y combinando los datos de forma posterior.

Esta característica es muy útil cuando se desean añadir muestras al análisis, como es nuestro caso. Para la anotación de las sondas y los genes hemos usado la versión Ensembl 63 del genoma, que corresponde al *Assembly* GRCh37.

3.3.2. Determinación de ausencia o presencia de transcritos (*expression calls*)

A continuación, hemos utilizado el método *Gene Expression Barcode* (al que denominamos *Barcode*), incluido en el paquete fRMA de R, y que convierte los valores de intensidad normalizados de cada gen en valores binarios (1 ó 0, refiriéndose a expresado o no expresado). La mayor ventaja de esta aproximación radica en su capacidad de minimizar los efectos de lote o *batch* (como pueden ser: diferentes tiempos de análisis, diferentes plataformas de hibridación, diferentes laboratorios...) y reducir el ruido [82].

3.3.3. Generación de espacios de expresión mediante MCA

En el siguiente paso, se generan los espacios de expresión génica, que permiten la visualización espacial y la clasificación en grupos de los patrones de expresión de las diversas muestras. Dado que no estamos utilizando las intensidades de expresión, sino los valores discretos de expresión o no expresión, usamos un análisis de correspondencias múltiples (MCA, de sus siglas en inglés *Multiple Correspondence Analysis*) [32]. MCA se puede considerar un método equivalente a PCA (*Principal Component Analysis*), pero más adecuado cuando se trabaja con valores cualitativos (o binarios) en lugar de variables continuas. Ambas son técnicas utilizadas para reducir la dimensionalidad de un conjunto de datos y determinar las causas de la variabilidad. En este trabajo, hemos adaptado un protocolo metodológico basado en MCA y utilizado previamente en alineamientos múltiples de secuencias para la detección de subfamilias de proteínas [93].

Nuestro protocolo comienza utilizando MCA para realizar una representación vectorial de los datos binarios, previamente sometidos al método *Barcode*, de los perfiles de expresión en un espacio multidimensional. Dicho espacio, es producido por el MCA de forma que las dimensiones están ordenadas en función del porcentaje de varianza explicado de los datos originales. Esto permite reducir el número de dimensiones a estudiar (las primeras) preservando la mayor parte de la información original. Esta reducción elimina las contribuciones menos relevantes que corresponden, típicamente, a las principales fuentes de ruido. Nuestro protocolo establece que el espacio dimensional adecuado para el análisis, es aquel compuesto por aquellas dimensiones cuya incorporación resulta en un incremento estadísticamente significativo de la información. El número de dimensiones informativas (aquellas que explican la mayor parte de la varianza total) se seleccionan según el valor p (*p-value*) de un análisis Wilcoxon entre ejes contiguos (en orden decreciente de porcentaje de varianza explicado). Se añade una nueva dimensión al espacio de expresión si p es inferior a 0.01. A continuación, para un número de grupos predeterminado (entre 2 y 50), se realiza un análisis de agrupamiento o *clustering* robusto no supervisado de *k-means* (este método agrupa las observaciones en k grupos, de modo que cada observación pertenece al grupo más cercano a la media) de forma iterativa en ese espacio reducido (definido por las dimensiones informativas). Finalmente, de todas las soluciones proporcionadas por el último paso, se consideran óptimas, y por tanto válidas, aquellas soluciones que, habiéndose obtenido en al menos un 10% de las iteraciones, maximizan el valor del índice CH (ICH. Índice Calinsky y Harabsz, descrito en 1974, que analiza la calidad del *clustering*) [93].

3.3.4. Análisis complementarios al MCA

Para apoyar los resultados obtenidos en los espacios de expresión génica obtenidos mediante MCA, llevamos a cabo análisis complementarios mediante dos técnicas. La primera es un análisis alternativo con una herramienta equivalente, PCA, usando los valores de intensidad normalizados. Utilizamos la función *prcomp* de R (paquete *stats*, versión 3.0.1.).

La segunda es una técnica más simple de análisis de agrupamiento o *clustering* jerárquico. Este análisis se lleva a cabo mediante la función *dist* de R (paquete *stats*, versión 3.0.1.). Se pueden seleccionar diferentes tipos de distancias, en función del tipo de dato de entrada. Nosotros utilizamos la distancia binaria (*binary*), debido a que es la recomendada cuando se utilizan valores binarios, en nuestro caso obtenidos mediante el método *Barcode*. Adicionalmente, usamos la distancia euclídea (*euclidean*) cuando los datos de entrada son los valores de las coordenadas obtenidos con el MCA.

3.4. Cuantificación del componente murino en los perfiles de expresión de modelos PDX

Las muestras usadas en esta tesis, incluyendo las muestras de modelos PDX, han sido hibridadas en una plataforma diseñada para muestras humanas (*GeneChip® Human Genome U133 Plus 2.0 Array*). Dichas muestras contienen parte de tejido murino, principalmente, debido a que el estroma humano es sustituido por tejido estromal del ratón. La plataforma de hibridación usada en esta tesis contiene 54675 sondas (*probesets*), que están constituidas por 20 oligonucleótidos diferentes (de 25 nucleótidos cada uno). Para comprobar si los perfiles de expresión estaban siendo contaminados por el tejido murino, hemos cuantificado cuantas secuencias de la plataforma de hibridación pueden considerarse ambiguas (podrían hibridar con tejido de ratón específicamente o con tejido humano o murino indistintamente). Para llevar a cabo el análisis hemos usado la herramienta *Xenome* [17]. Dicha herramienta nos permite analizar detalladamente la secuencia de los oligonucleótidos (1093500) contenidos en la plataforma, frente a todos los transcritos anotados en el genoma humano (*Ensembl Assembly GRCh37*) y de ratón (*Ensembl Assembly GRCm38*). Finalmente, nos proporciona el porcentaje de similitud máximo que encuentra en cada uno de los dos genomas.

3.5. Cuantificación del componente no tumoral

Para determinar la cantidad de tejido estromal presente en una muestra de tumor primario y compararla con la de un modelo PDX (generado a partir del mismo paciente), hemos realizado una tinción tricrómica de Masson. Esta tinción utiliza tres colorantes (escarlata de Biebrich o fucsina ácida, ácido fosfomolibdico y azul de anilina) para diferenciar el núcleo celular, el citoplasma y las fibras de colágeno (colágeno tipo I principalmente). Esta técnica nos permite visualizar claramente las áreas fibróticas de una muestra. Las fibras de colágeno se tiñen de azul, el citoplasma de color entre lila y rojo y, por último, el núcleo se tiñe de color entre rojo y marrón.

En nuestro caso, la muestra de tumor primario estaba congelada, mientras que la del modelo PDX estaba en un bloque de parafina. Por ello, las tinciones muestran diferencias a nivel visual. Para facilitar la interpretación hemos realizado una cuantificación del tejido estromal presente. Dicha cuantificación se lleva a cabo estableciendo el ratio área azul/área coloreada, es decir, área fibrótica/área tisular total. Usamos la zona coloreada total para eliminar de la cuantificación las zonas necróticas o zonas dañadas por el proceso de congelación.

3.6. Estimación y corrección del componente no tumoral

Hemos recopilado una gran cantidad de perfiles de expresión, pero no disponemos de las muestras correspondientes. Por tanto, no podemos hacer una cuantificación de la cantidad de estroma de las mismas, sin embargo podemos llevar a cabo una estimación. La estimación de la proporción de células estromales e inmunes infiltradas en las muestras se ha llevado a cabo mediante ESTIMATE (*Estimation of STromal and Immune cells in MAlignant Tumours using Expression data*, versión 1.0.11) [127]. ESTIMATE es un método capaz de establecer la contribución de las células no tumorales, pertenecientes al estroma y al sistema inmune, a un perfil de expresión. El método está basado en la expresión de dos firmas génicas, una que contiene genes representantes del estroma y otra con genes típicos de las células

inmunes infiltrantes. En este estudio, ambas firmas se han combinado dando lugar a una única firma, la firma de las células no tumorales infiltrantes (NTI). El algoritmo ha sido ejecutado usando el código original (en lenguaje de programación R, versión 3.0.3), proporcionado por los autores y con los valores por defecto para la plataforma Affymetrix.

A continuación, hemos intentado eliminar (*in silico*) el componente no tumoral de los espacios de expresión. Para ello, primero se han generado espacios de expresión génica alternativos. Los espacios alternativos se obtienen eliminando del análisis las sondas que corresponden a los genes de la firma de las células NTI. Adicionalmente, sobre dichos espacios se elimina la señal de la expresión directamente relacionada con las células NTI, mediante modelos de regresión lineal simple. Se establece un modelo de regresión para cada uno de los componentes del espacio que muestren una correlación significativa con los valores establecidos por ESTIMATE (usados como variable independiente). Los residuos de cada regresión se utilizan para generar el nuevo espacio de expresión génica “corregido” de acuerdo a la contribución de las células NTI. Es importante tener en cuenta que no es una corrección total, dado que para que así fuera debería aplicarse dicha corrección sobre los datos normalizados (antes de crear el espacio) y la nuestra se realiza sobre un espacio de expresión génica previamente creado. La eliminación y corrección completa del efecto de las células del estroma y sistema inmune nos parece un idea muy interesante, pero debido a que no es una cuestión trivial y fácil de solucionar, todavía estamos trabajando en ello.

3.7. Análisis de enriquecimiento funcional

3.7.1. Análisis basado en frecuencias de expresión

Para estudiar las diferencias entre los tumores originales de los pacientes y los modelos PDX, nos centramos en su función y, por tanto, en las rutas biológicas. Primero calculamos el número de veces que cada gen se considera expresado en cada grupo, es decir, calculamos su frecuencia de expresión. Consideramos que un gen se

expresa si alguna de las sondas presentes en la plataforma de hibridación, y cuya localización corresponde a ese gen, es considerada como expresada según los criterios del método *Barcode*. A continuación, calculamos la diferencia de frecuencia de expresión de cada gen en los modelos PDX respecto a los tumores primarios. En el siguiente paso, ordenamos los valores de la diferencia de frecuencia de expresión y realizamos un análisis de enriquecimiento de conjuntos de genes (*Gene Set Enrichment Analysis*, GSEA), con la herramienta *GseaPreranked* del programa GSEA [111]. GSEA es una herramienta informática capaz de proporcionar medidas estadísticas de la acumulación de genes sobre-expresados en un conjunto de genes determinado, cuando comparamos dos fenotipos diferentes (por ejemplo, tumores primarios y modelos PDX). Como conjuntos de genes utilizamos las 1307 rutas biológicas anotadas en Reactome [78]. Los filtros presentes por defecto en la aplicación (mínimo 15 y máximo 500 genes en cada grupo), provocan que el grupo final de rutas analizadas conste de 621. Aquellas rutas biológicas consideradas significativamente enriquecidas ($FDR < 0.05$) las agrupamos en “grupos funcionales”, según la jerarquía de Reactome.

3.7.2. Análisis basado en t-test sobre las intensidades

Cuando el número de muestras es bajo, no podemos calcular frecuencias de expresión, por lo que en estas situaciones usamos métodos alternativos basados en el t-test. En nuestro caso, para comprobar si los modelos PDX son similares a las metástasis, estudiamos las alteraciones funcionales, analizando si los cambios presentes en los modelos PDX también se encuentran alterados en las muestras de metástasis.

Primero, comprobamos cuales son los genes diferencialmente expresados en las metástasis respecto a los tumores primarios. Para ello, utilizamos los valores normalizados de expresión como datos de entrada. Después, los genes se ordenan según la métrica t-test. A continuación, se analiza si existe una acumulación de los genes diferencialmente expresados en metástasis en los grupos funcionales (significativamente enriquecidos en modelos PDX).

3.8. Expresión diferencial entre diferentes pases

Para estudiar si el número de pases influye en los cambios presentes en los modelos PDX respecto de los tumores primarios, hemos usado un conjunto de datos adicional. Éste contiene los perfiles de expresión de 4 pacientes diferentes, hibridados en la misma plataforma que hemos usado en todo este trabajo (*GeneChip® Human Genome U133 Plus 2.0 Array*). Para cada paciente disponemos del tumor primario (F0), y de los modelos PDX en pase 5 (F5) y pase 10 (F10) [30]. Para testear la expresión diferencial de los genes entre los 4 tumores de los pacientes y sus correspondientes modelos PDX, usamos el paquete *limma* de R (versión 3.16.8) [104]. Primero creamos un modelo lineal con la información del número de pase (F0, F5 y F10), usando la pertenencia a cada paciente como información de bloque. Entonces realizamos varios análisis de contraste, F0 versus F5, F5 versus F10 y F0 versus F10. Tras cada análisis, ordenamos los genes según el valor *t* y usamos la herramienta *GseaPreranked*. Los resultados nos permiten analizar si los niveles de expresión de los genes, cambian en los modelos PDX respecto de los tumores primarios originales, así como si se producen cambios entre los pases.

3.9. Estudio de eficacia de un fármaco

Un estudio de eficacia permite evaluar si los pacientes incluidos en él son sensibles a un determinado fármaco. En nuestro caso, el objeto de estudio son los modelos PDX, lo que nos permite determinar que fármaco es más eficaz en cada paciente. Para llevar a cabo un estudio de eficacia, se diseña un experimento con tantas ramas de tratamiento como fármacos se quieran analizar, y además un rama control. Los tumores se implantan en los ratones y se dejan crecer hasta alcanzar un volumen medio de 200 mm³. Cuando los tumores alcanzan ese valor, los modelos PDX se distribuyen aleatoriamente en los grupos de tratamiento (manteniendo el valor medio de 200 mm³ en cada grupo). Para valorar la eficacia de cada fármaco, usamos el factor de inhibición del crecimiento tumoral (TGI, *Tumor Growth Inhibition*). $TGI = 1 - [(T_f - T_i) / (C_f - C_i)]$, siendo T = tumor, C = control, f = final, i = inicial).

Para esta tesis, diseñamos un experimento incluyendo los modelos PDX, tanto derivados de tumor primario como de metástasis, de los pacientes Panc042 y Panc047. El experimento consta de tres ramas de tratamiento y la rama control (suero salino, diariamente, mediante administración intraperitoneal). Los tres fármacos seleccionados fueron Gemcitabina (100 mg/kg, dos veces por semana, mediante administración intraperitoneal), Mitomicina C (5 mg/kg, una única dosis, mediante administración intraperitoneal) y Erlotinib (50 mg/kg, diariamente, mediante administración oral).

3.10. Predicción de los reguladores transcripcionales asociados a los modelos PDX

A lo largo de esta tesis, hemos usado perfiles de expresión de diversas muestras: líneas celulares, tumores primarios completos y microdisecccionados, modelos PDX derivados del tumor primario (a veces procedentes de diferentes países), metástasis frescas... Es importante tener en cuenta que disponemos de perfiles de expresión de tumor primario, metástasis, modelo PDX derivado de tumor primario y modelo PDX derivado de metástasis de 2 pacientes, Panc042 y Panc047. Nuestro siguiente paso es estudiar que diferencias o similitudes podemos rescatar entre las muestras de un mismo individuo. Para ello, planteamos los perfiles de expresión de dichas muestras como “procesos evolutivos” (por ejemplo: primero se produce el tumor primario en el paciente, evoluciona a metástasis y, finalmente, lo implantamos en un ratón generando un modelo PDX derivado de metástasis). Para analizar esas diferencias y similitudes, hemos comenzado estudiando la activación o desactivación de los sitios de unión a factores de transcripción (TFBS, *Transcription Factor Binding Sites*) y los posibles reguladores transcripcionales asociados.

Para analizar el estado de dichos motivos hemos usado el programa ISMARA (*Integrated System for Motif Activity Response Analysis*) [4]. ISMARA utiliza el modelo MARA (*Motif Activity Response Analysis*) [112], el cual calcula la actividad de una muestra mediante la asociación de los niveles de expresión a la presencia de

los promotores. El programa identifica los promotores y calcula la señal transcripcional asociada a ellos en cada muestra. A continuación, estudia posibles patrones de regulación en dicha señal, para una lista predeterminada de motivos o TFBS conocidos y cercanos a los promotores. ISMARA finalmente predice que interacciones pueden estar regulando los TFBS. El programa, además, permite reanalizar los resultados teniendo en cuenta si parte de las muestras presentan algún tipo de efecto de lote o *batch*. Nosotros hemos aplicado esa aproximación cuando las muestras lo permitían. ISMARA proporciona un valor z que permite evaluar la significancia del análisis. Los autores recomiendan no usar resultados con un valor inferior a 2, por lo que para ser un poco más restrictivos, nosotros solo hemos considerado significativos aquellos con un valor z , *z-score*, superior a 3.

4. Resultados

4.1. Modelos PDX derivados de tumor primario

4.1.1. Espacio de expresión génica de ADP

Para el estudio de la expresión génica en el ADP, hemos seleccionado los modelos PDX implantados subcutáneamente por diversos motivos, expuestos en la Introducción. El conjunto de datos de expresión de ADP analizado en esta tesis contiene un total de 147 muestras incluyendo: 53 modelos PDX, 53 muestras de tumor primario y 41 líneas celulares (Tabla S1, Anexo I – Material suplementario). Parte de los perfiles de expresión se han obtenido de los modelos PDX generados para el biobanco del CNIO, mientras que otros han sido obtenidos de bases de datos públicas. Todos los datos fueron procesados siguiendo el protocolo de análisis que hemos desarrollado para explorar cuantitativamente el espacio transcripcional de muestras biológicas, a partir de sus perfiles de expresión binarios (expresado o no expresado, ver Materiales y métodos para obtener más detalles).

Denominamos espacio de expresión génica a la distribución de las muestras, de acuerdo a sus perfiles de expresión, y que refleja las similitudes y diferencias entre las mismas. Para obtener estos espacios de expresión génica, utilizamos una metodología basada en MCA (*Multiple Correspondence Analysis*) [93]. Se representa cada muestra como un vector de los datos de expresión de cada sonda. Dichas muestras se organizan en un espacio de N-dimensiones (que por simplicidad representamos en muchos casos en solo dos o tres dimensiones), de modo que refleja del modo más exacto posible sus distancias. Las distancias entre muestras, obtenidas con esta aproximación, son posteriormente agrupadas en clases utilizando un método de agrupamiento o *clustering* que, adicionalmente, asigna un nivel de significación estadística a cada uno de estos grupos.

La Figura 7A muestra el espacio de expresión génica de cáncer de páncreas, generado con perfiles de expresión de modelos PDX, tumores primarios y líneas celulares.

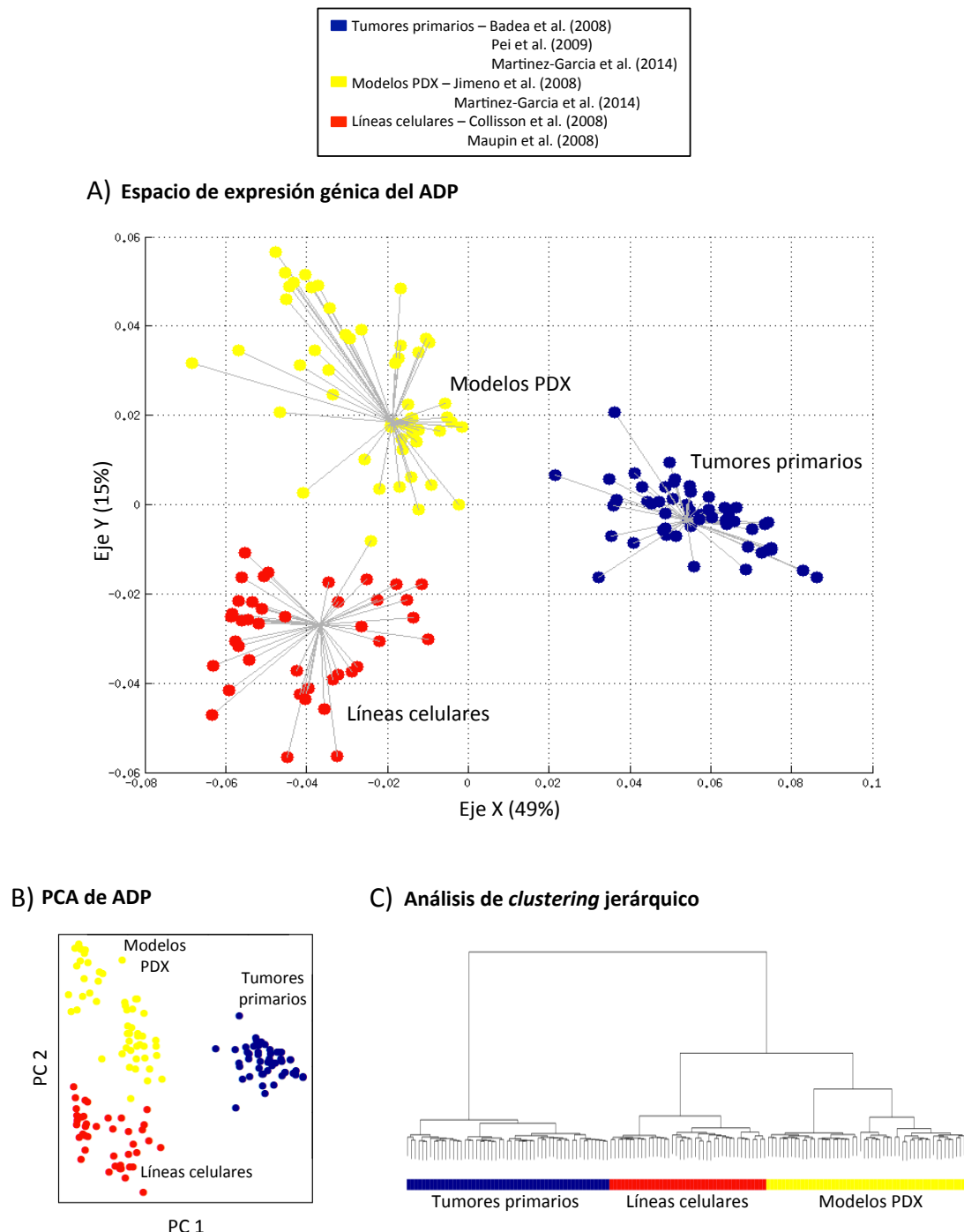


Figura 7. A) Espacio de expresión génica del ADP, obtenido mediante MCA de los datos de expresión binarios. Cada punto representa el patrón de expresión de una muestra. El primer eje de separación es el eje X, el segundo es el eje Y. Las líneas

grises representan los grupos resultantes del análisis de *clustering* mediante *k-means* para $k = 3$ (ICH = 199.54). Cada línea une una muestra con el centro de masas del grupo al que pertenece. **B)** Espacio de expresión complementario mediante PCA utilizando las intensidades normalizadas. **C)** Resultado complementario mediante análisis de *clustering* jerárquico de los valores binarios de expresión.

Inicialmente generamos un espacio multidimensional, el cual reducimos seleccionando las dimensiones que contienen la mayor parte de la información de partida. Para este espacio de expresión se seleccionan dos dimensiones, puesto que un Wilcoxon test entre ejes contiguos muestra que existen diferencias significativas solo entre el primer y segundo eje ($p = 2 \times 10^{-4}$, siendo el valor umbral de significancia 0.01). Sin embargo, la diferencia entre el segundo y tercero no es significativa ($p = 0.08$), lo que significa que el tercer eje añade muy poca información. El conjunto de las dos primeras dimensiones explica el 64% de la varianza (49% el primer eje, eje X, y 15% el segundo, eje Y. El tercer eje solo explica el 8%).

El MCA resultante muestra de forma clara que las muestras de tumor primario humano, los modelos PDX y las líneas celulares son tres grupos distintos (Figura 7A), si tenemos en cuenta los perfiles de expresión de sus genomas. A continuación, se lleva a cabo el análisis de *clustering*, mediante el método *k-means*, sobre el espacio creado por el MCA. Para medir la calidad del *clustering* y decidir si las soluciones encontradas son óptimas, usamos el índice CH (ICH). En este caso, el *clustering* encuentra una solución óptima correspondiente a 3 grupos ($k = 3$, ICH = 199.54): (1) tumores primarios originales, (2) modelos PDX y (3) líneas celulares. Resultados similares se obtuvieron al analizar las intensidades mediante PCA (*Principal Component Analysis*, ver Figura 7B) y mediante el análisis de *clustering* jerárquico de los valores binarios de expresión proporcionados por el método *Barcode* (Figura 7C). En resumen, estos resultados muestran que los modelos PDX y las líneas celulares son diferentes transcripcionalmente, en oposición a lo expuesto por algunos autores [28].

Para estudiar si existe una diferencia a nivel transcripcional entre los modelos PDX implantados de forma subcutánea y los modelos ortotópicos, hemos generado un nuevo espacio de expresión añadiendo modelos PDX ortotópicos procedentes de otro laboratorio [120] (Tabla S1, Anexo I – Material suplementario).

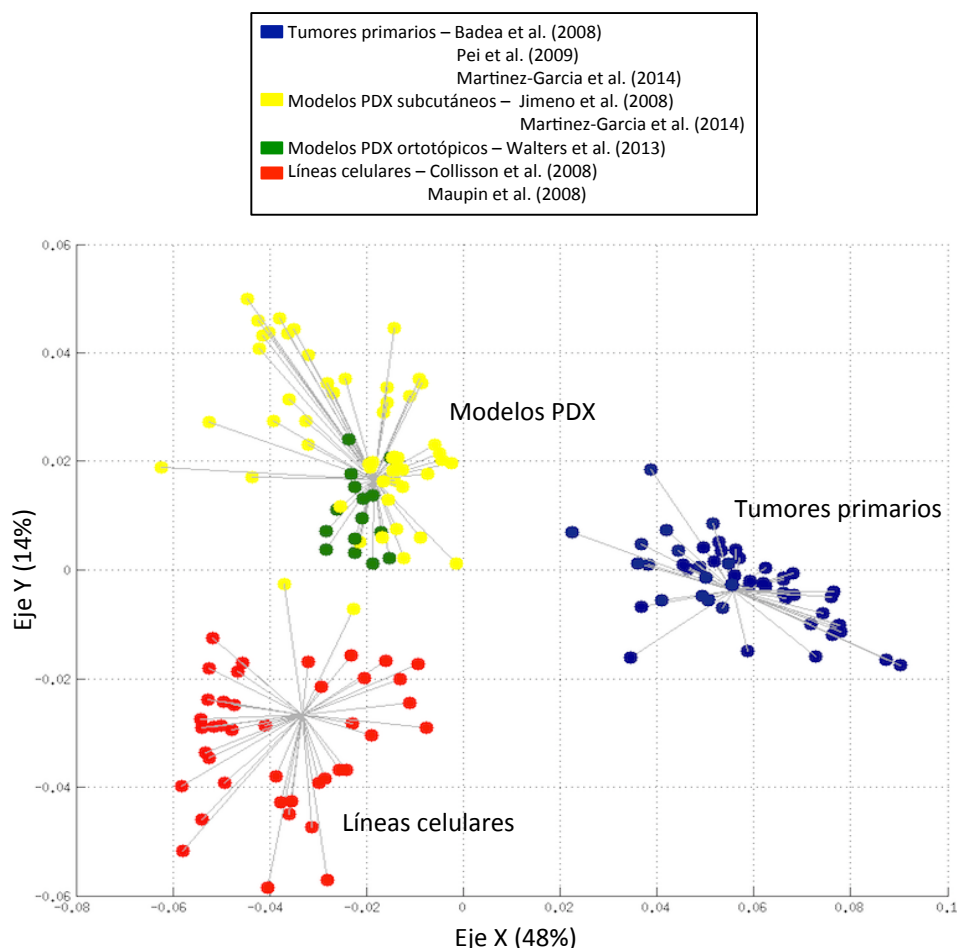


Figura 8. Espacio de expresión génica de ADP incluyendo tumores primarios, líneas celulares, modelos PDX implantados de forma subcutánea y ortotópica. Cada punto representa el patrón de expresión de una muestra. El primer eje de separación es el eje X, el segundo es el eje Y ($p = 9.6 \times 10^{-5}$). Las líneas grises representan los grupos resultantes del análisis de *clustering* mediante *k-means* para $k = 3$ (ICH = 221.92). Cada línea une una muestra con el centro de masas del grupo al que pertenece.

En la Figura 8, se puede observar que los modelos ortotópicos y subcutáneos son indistinguibles a nivel transcripcional. Además, es importante tener en cuenta que la distribución del resto de muestras en el espacio se mantiene prácticamente idéntica (comparar Figuras 7A y 8A). A partir de este punto, utilizaremos solo los datos correspondientes a los modelos PDX subcutáneos, ya que son con los que trabajamos de rutina en nuestro laboratorio.

4.1.2. Espacio de expresión génica de carcinoma hepatocelular (CHC)

Para realizar una comparación de los resultados obtenidos con las muestras de ADP con otros tipos de cáncer, realizamos una búsqueda exhaustiva de datos de expresión en la misma plataforma de hibridación que hemos utilizado previamente, intentando reunir un número suficiente de muestras para los distintos grupos experimentales (tumor primario, líneas celulares y modelos PDX). Pudimos recopilar suficientes datos únicamente para el caso del carcinoma hepatocelular (CHC). El conjunto de datos de CHC consta de 182 muestras (Tabla S1, Anexo I – Material suplementario). Para realizar el análisis de estos datos, aplicamos el mismo protocolo usado previamente con los datos de ADP. La Figura 9A muestra el espacio transcripcional de CHC basado en las dos primeras dimensiones del MCA, las cuales explican el 57% de la varianza (41% el primer eje, eje X, y 16% el segundo, eje Y; $p = 1.6 \times 10^{-7}$).

Los modelos PDX de CHC y las líneas celulares se localizan en diferentes áreas del espacio (Figura 9A). El agrupamiento por *k-means* del espacio encuentra una solución óptima de tres grupos ($k = 3$, $ICH = 236.30$): (1) tumores primarios humanos, (2) modelos PDX y (3) líneas celulares. La comparación de los resultados del MCA respecto a los obtenidos mediante PCA (realizado con las intensidades normalizadas), el cual, se muestra en la Figura 9B y el *clustering* jerárquico (utilizando los valores binarios de expresión), que se puede observar en la Figura 9C, muestran agrupamientos similares. En este caso, podemos observar que se obtienen resultados equivalentes a los obtenidos con el espacio de expresión génica de ADP, es decir, que las líneas celulares, los modelos PDX y los tumores primarios son tres grupos independientes.

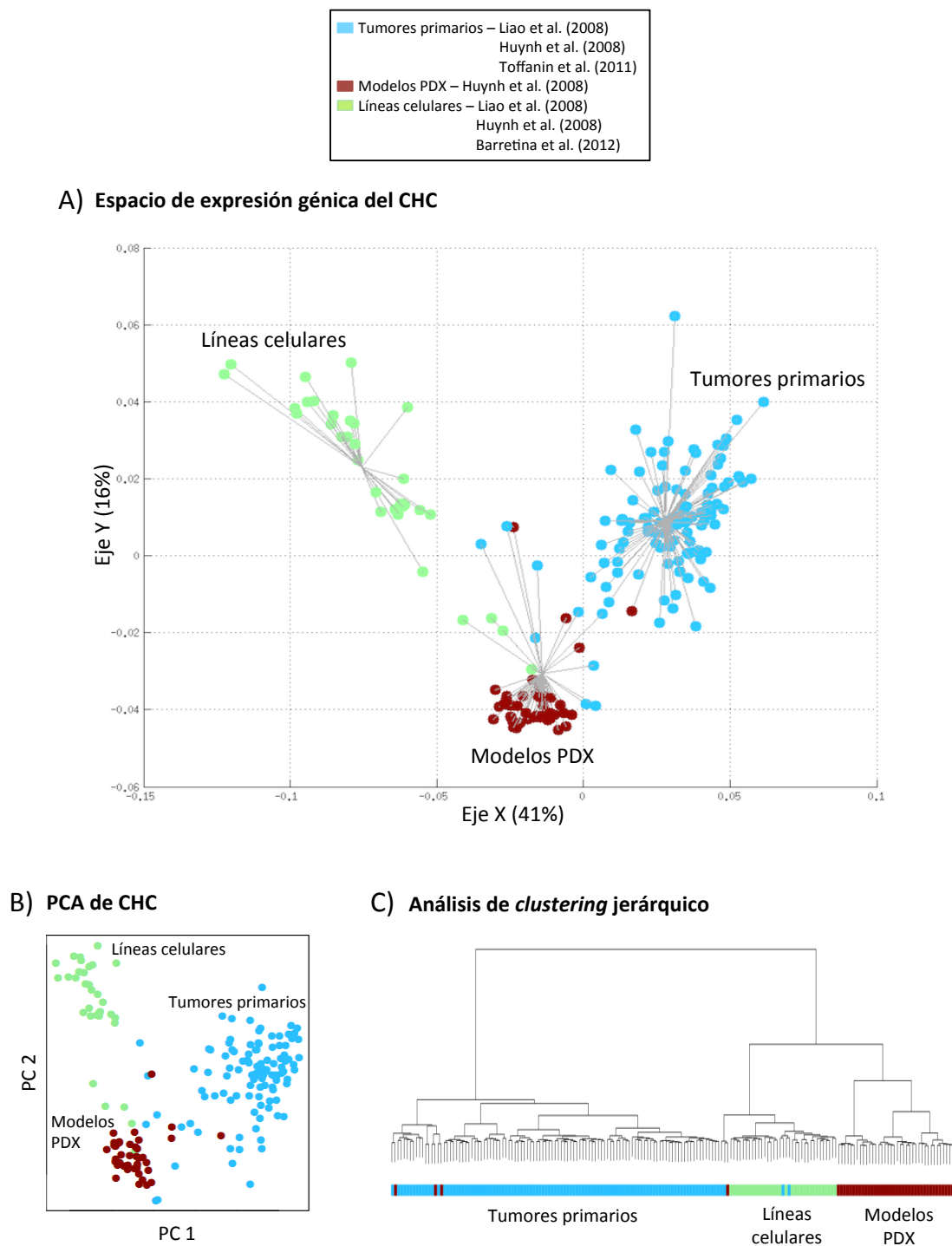


Figura 9. **A)** Espacio de expresión génica de CHC, obtenido mediante MCA de los datos de expresión binarios. Cada punto representa el patrón de expresión de una muestra. El primer eje de separación es el eje X, el segundo es el eje Y. Las líneas grises representan los grupos resultantes del análisis de *clustering* mediante *k-means* para $k = 3$ ($ICH = 236.30$). Cada línea une una muestra con el centro de masas del grupo al que pertenece. **B)** Espacio de expresión complementario mediante PCA

utilizando las intensidades normalizadas. C) Resultado complementario mediante análisis de *clustering* jerárquico de los valores binarios de expresión.

4.1.3. Espacios de expresión génica combinando muestras de ADP y CHC

Es razonable pensar que las células tumorales humanas injertadas en ratón pueden conservar una firma específica de expresión de su tejido de origen. Sin embargo, si el microambiente murino induce cambios importantes, entonces las células adquirirían unas características injerto-específicas que podrían conllevar, en último término, a perder totalmente la huella de expresión original.

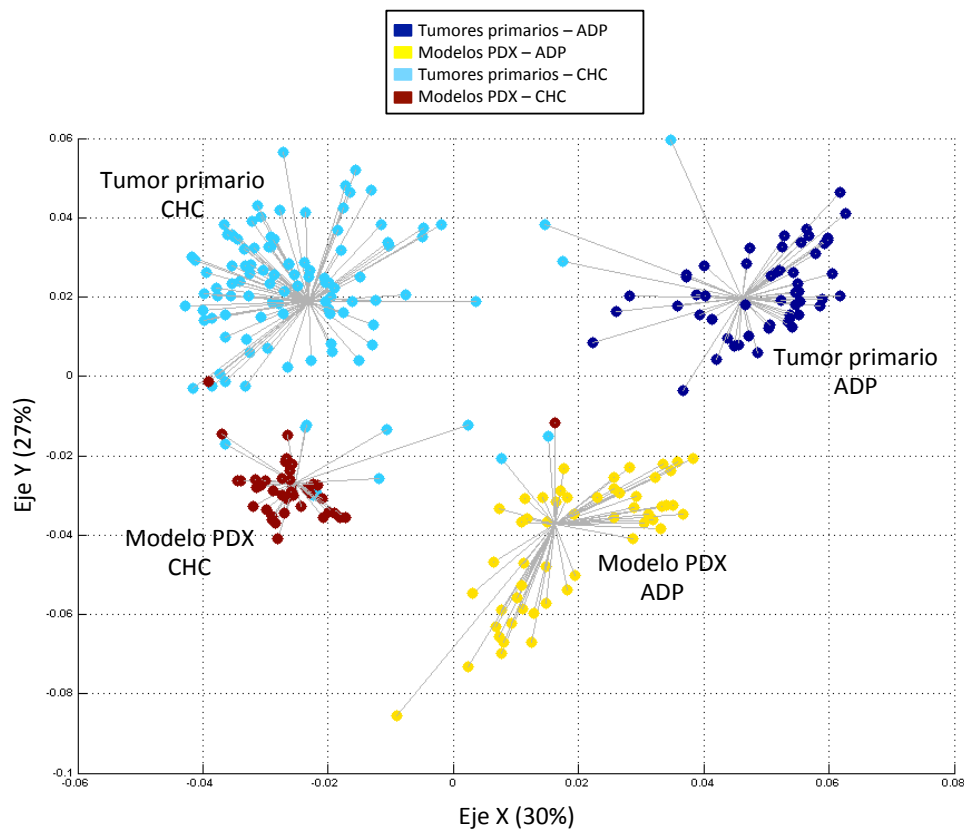


Figura 10. Espacio de expresión génica de tumores primarios y modelos PDX, obtenido mediante MCA de los datos de expresión binarios. Cada punto representa el patrón de expresión de una muestra. Las líneas grises representan los grupos resultantes del análisis de agrupamiento para $k = 4$ ($ICH = 228.70$). Cada línea une una muestra con el centro de masas del grupo al que pertenece.

Para estudiar el efecto del ambiente murino sobre el tumor injertado, hemos analizado simultáneamente perfiles de expresión de tumores primarios y modelos PDX de los dos tipos tumorales presentes en nuestros datos, ADP y CHC. La Figura 10 muestra el espacio de expresión resultante.

El primer componente del espacio obtenido (Figura 10, eje X), explica el 30% de la varianza y separa las muestras según su tejido de origen, es decir, separa los tumores primarios y los modelos PDX de páncreas de los tumores primarios y los modelos PDX de CHC, los cuales agrupan por separado. El segundo componente (Figura 10, eje Y) separa los tumores primarios de los modelos PDX y explica el 27% de la varianza. Podríamos decir que rescatamos dos efectos, uno que representa la conservación de una memoria de expresión en los modelos PDX respecto a los tumores primarios, y un segundo efecto que representa la diferencia del huésped o ambiente.

En este caso, a diferencia de los espacios de expresión génica anteriores, son tres los ejes seleccionados como informativos, debido a que la información que añade el tercer eje es significativa según el test de Wilcoxon (entre los ejes X e Y, $p = 4.04 \times 10^{-29}$; entre los ejes Y y Z, $p = 6 \times 10^{-4}$). En la Figura 11 se puede observar una visión tridimensional del espacio de expresión resultante, y las combinaciones bidimensionales de los tres ejes. El tercer componente, correspondiente al eje Z, explica el 8% de la varianza. La incorporación de este eje al espacio, como se puede ver en la Figura 11, contribuye a la mejor separación de los cuatro grupos.

La solución óptima, obtenida según el *clustering* por *k-means*, para este espacio de expresión génica es de cuatro grupos ($k = 4$, ICH = 228.70) y clasifica por separado las muestras de tumor primario de ADP, los modelos PDX de ADP, los tumores primarios de CHC y los modelos PDX de CHC.

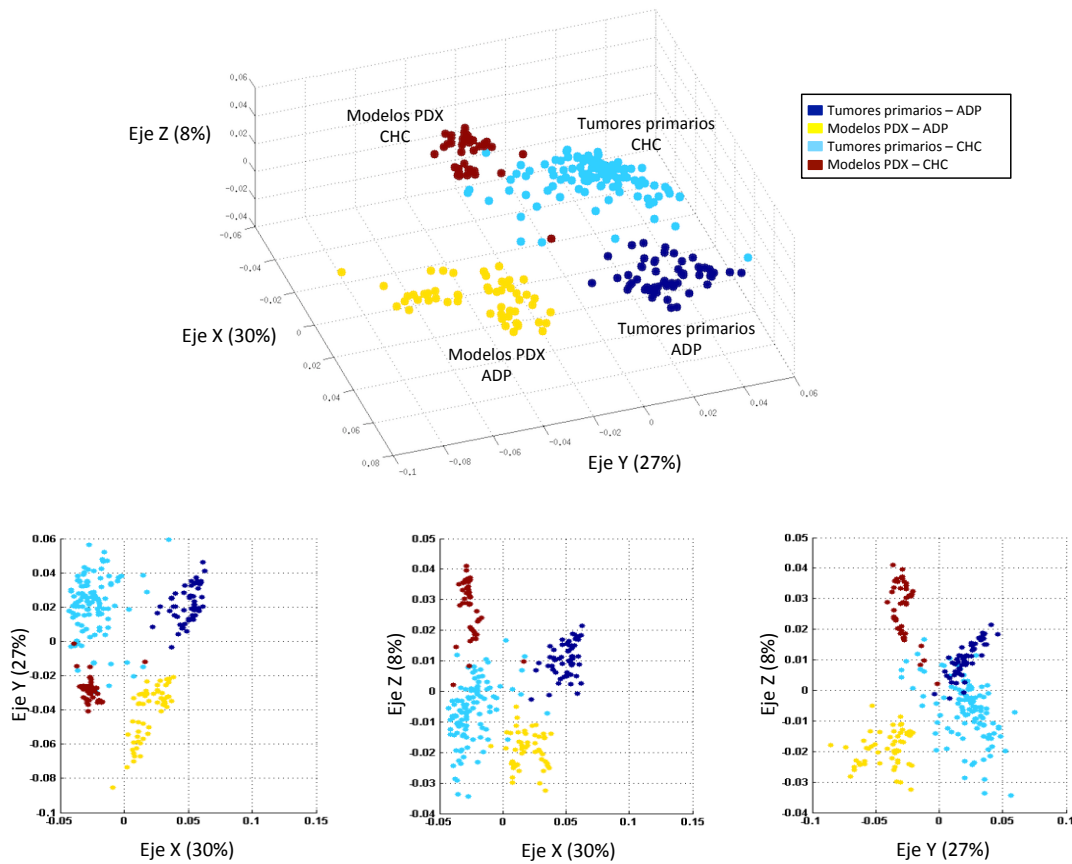


Figura 11. Visión tridimensional del espacio de expresión génica de tumores primarios y modelos PDX, de ADP y CHC, obtenido mediante MCA de los datos de expresión binarios. Cada punto representa el patrón de expresión de una muestra. En la parte superior de la figura se puede observar una visión tridimensional del espacio, y en la parte de abajo se encuentran las combinaciones de los tres ejes representados de forma bidimensional. Solución óptima para $k = 4$ (ICH = 228.70).

Es importante tener en cuenta, que el protocolo que hemos desarrollado lleva a cabo la asignación de los grupos (*k-means*) de manera no supervisada, es decir, no se tiene en cuenta de que tipo es cada muestra al realizar la clasificación en el espacio (ver sección Materiales y métodos). Por ello, podemos comprobar cuantas muestras se clasifican correctamente según el protocolo automático de selección de grupos. En páncreas, el 100% de los tumores primarios y los modelos PDX se agrupan en sus respectivas clases, mientras que en CHC el 88% de los tumores primarios y el 95% de los modelos PDX clasifican correctamente. En total, el 94% del total de 256 muestras, usadas para generar este espacio, es clasificado de manera correcta.

Una vez más, nuestros resultados usando los espacios de expresión génica basados en MCA, están apoyados por resultados similares obtenidos mediante análisis adicionales: PCA usando las intensidades de expresión (Figura 12A) y un análisis más simple de *clustering* jerárquico (Figura 12B-E) utilizando los valores binarios. En el caso de los análisis de *clustering* jerárquico, cuando utilizamos los valores discretos proporcionados por *Barcode* y la distancia binaria (Figura 12B), la clasificación es algo diferente a la obtenida mediante la distancia euclídea de las dimensiones del MCA (Figura 12C). Esto se debe a que el *clustering* jerárquico, obtenido usando la distancia binaria, utiliza toda la información que se distribuye en las dimensiones del espacio MCA (sin descomposición de las fuentes de variación). Cuando calculamos la distancia euclídea a partir de las dimensiones del MCA, solo consideramos los ejes informativos y estamos, por tanto, filtrando las principales fuentes de ruido (las cuales no se incluyen en el análisis).

Es muy importante tener en cuenta que el MCA considera cada eje como una fuente independiente de información y, por ello, cada eje puede rescatar una fuente de variabilidad diferente. De hecho, como ya hemos mencionado, en la Figura 10 podemos ver que el primer eje separa en función del tejido de origen, mientras que el segundo componente agrupa las muestras según el huésped. Ambos efectos son rescatados cuando realizamos los análisis de *clustering* jerárquico, usando como métrica para calcular la distancia (euclídea) entre muestras las coordenadas de cada dimensión del MCA de forma independiente. Primero, seleccionamos la coordenada del primer componente, eje X, que separa las muestras en función del tejido de origen en el MCA. En la Figura 12D, podemos ver claramente que la distribución de muestras coincide con una separación según el tejido de origen, agrupando las muestras derivadas de ADP y CHC por separado, y corroborando que existe una parte de los perfiles de expresión que se mantiene en los modelos PDX respecto a los tumores primarios. A continuación, seleccionamos la coordenada del segundo componente, el eje Y, que separa las muestras según el huésped. La Figura 12E, muestra que los perfiles de expresión se agrupan según el tipo de huésped, es decir, humanas (tumores primarios) o murinas (modelos PDX).

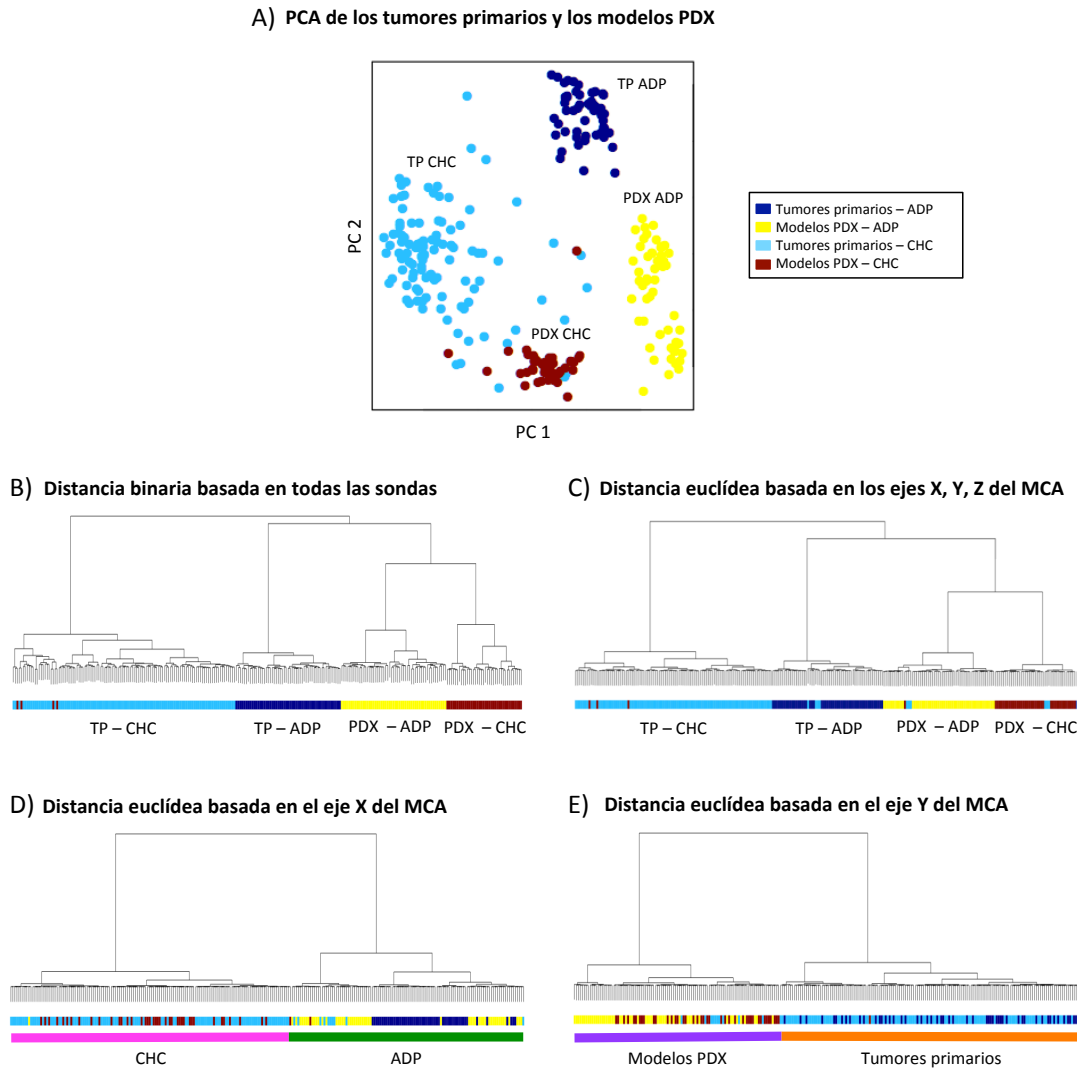


Figura 12. Análisis complementarios al espacio de expresión génica de tumores primarios y modelos PDX (Figura 10). **A)** Espacio de expresión obtenido mediante PCA utilizando las intensidades normalizadas. **B)** Análisis de *clustering* jerárquico utilizando la distancia binaria de todas las sondas. **C)** Análisis de *clustering* jerárquico utilizando la distancia euclídea calculada a partir de las coordenadas informativas (X, Y, Z) del espacio MCA. **D)** Análisis de *clustering* jerárquico utilizando la distancia euclídea calculada a partir de la coordenada X del espacio MCA. **E)** Análisis de *clustering* jerárquico utilizando la distancia euclídea calculada a partir de la coordenada Y del espacio MCA.

Nuestra interpretación de estos resultados es que el transcriptoma original de los tumores primarios humanos se ve alterado por el ambiente murino, pero conservando características específicas de su tumor de origen. Para comprobar si este efecto se observa también en las líneas celulares derivadas de tumores, realizamos un análisis similar usando las líneas celulares en lugar de los modelos PDX. La Figura 13 muestra el espacio de expresión génica resultante al utilizar los perfiles de expresión de los tumores primarios y las líneas celulares de ADP y CHC. Las dos dimensiones, añadidas al espacio por considerarse informativas, explican el 65% de la varianza (43% el primer eje, eje X, y 22% el segundo, eje Y; $p = 3.99 \times 10^{-13}$. La tercera dimensión no es significativa, $p = 0.14$).

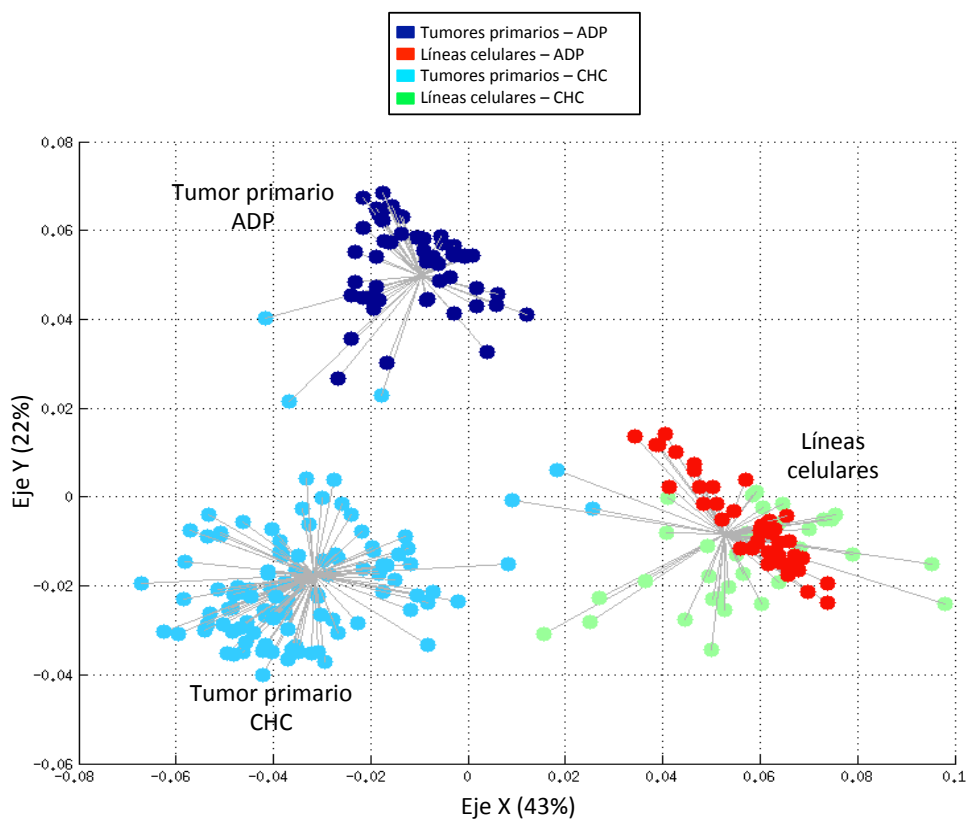


Figura 13. Espacio de expresión génica de tumores primarios y líneas celulares, obtenido mediante MCA de los datos de expresión binarios. Cada punto representa el patrón de expresión de una muestra. El primer eje de separación es el eje X, el segundo el eje Y. Las líneas grises representan los grupos resultantes del análisis de *clustering* para $k = 3$ ($ICH = 373.85$). Cada línea une una muestra con el centro de masas del grupo al que pertenece.

Es de remarcar que mientras los modelos PDX derivados de ADP y CHC formaban dos grupos distintos (Figura 10), las correspondientes líneas celulares son indistinguibles. Como en los análisis previos, también se han llevado a cabo análisis complementarios con técnicas de PCA con las intensidades normalizadas (Figura 14A) y de *clustering* jerárquico con los valores binarios (Figura 14B-E).

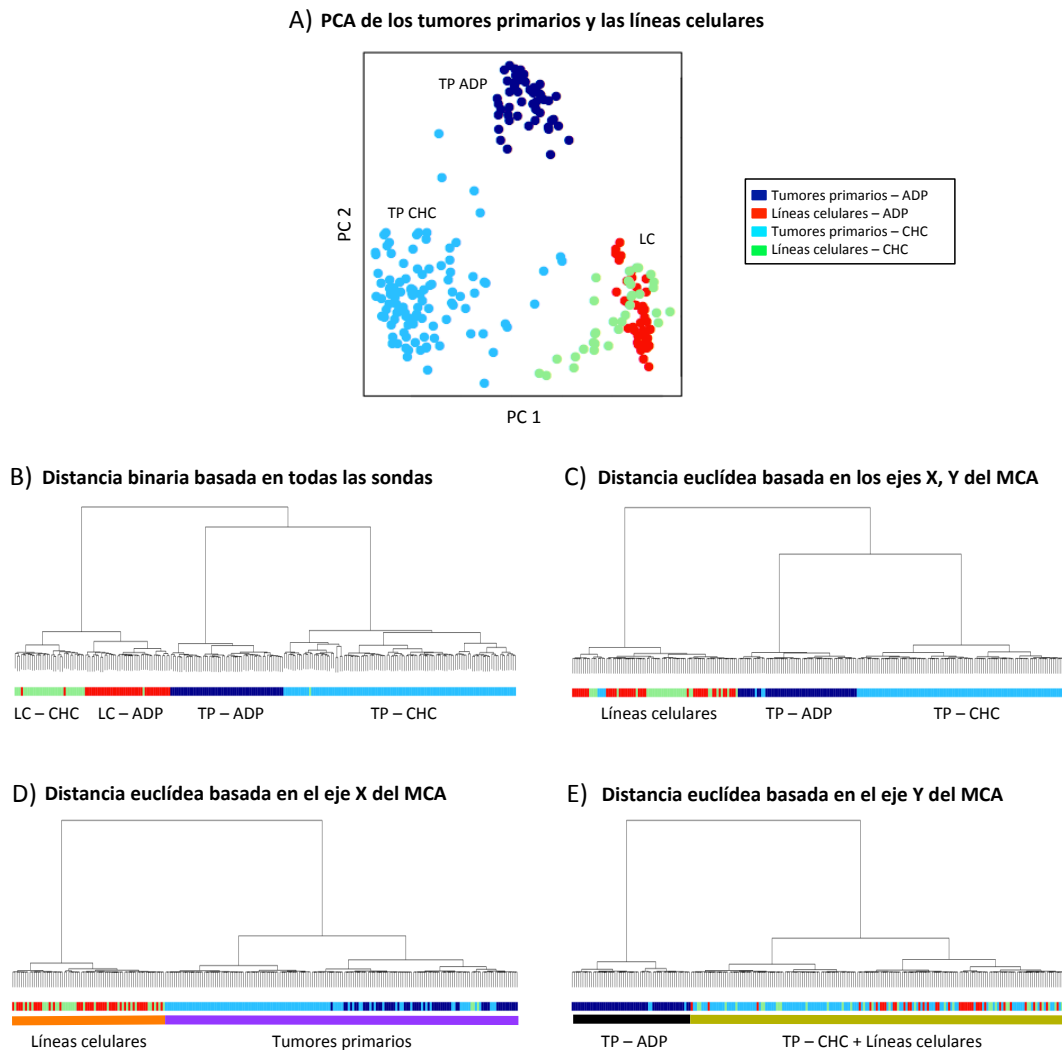


Figura 14. Análisis complementarios al espacio de expresión génica de tumores primarios y líneas celulares (Figura 13). **A)** Espacio de expresión obtenido mediante PCA utilizando las intensidades normalizadas. **B)** Análisis de *clustering* jerárquico utilizando la distancia binaria de todas las sondas. **C)** Análisis de *clustering* jerárquico utilizando la distancia euclídea calculada a partir de las coordenadas informativas, X e Y, del espacio MCA. **D)** Análisis de *clustering* jerárquico utilizando la distancia

euclídea calculada a partir de la coordenada X del espacio MCA. E) Análisis de *clustering* jerárquico utilizando la distancia euclídea calculada a partir de la coordenada Y del espacio MCA.

Al igual que los análisis de *clustering* jerárquico de los tumores primarios y los modelos PDX, en este caso, el agrupamiento es diferente al considerar los ejes del MCA de forma independiente. La Figura 14D muestra que el primer componente de separación es debido al efecto huésped, ya que agrupan por separado los tumores primarios y las líneas celulares. Debido a que solo hay tres grupos, y que las líneas celulares son indistinguibles entre ellas, el segundo componente muestra una agrupación que no podemos asignar a una fuente de variabilidad determinada. En resumen, detectamos una similitud en la expresión génica entre tumores primarios y modelos PDX que parece haberse perdido en las líneas celulares.

4.1.4. Contribución de las células no tumorales infiltrantes al perfil de expresión de los modelos PDX

La sustitución del estroma original humano de los tumores primarios por el estroma murino en los modelos PDX es un factor importante que puede contribuir a las diferencias observadas en los perfiles de expresión de ambos modelos. La Figura 15A, permite ilustrar esa sustitución. La imagen ha sido cedida por el Dr. López-Casas (datos no publicados), y muestra un modelo PDX en el cual se ha llevado a cabo la detección de secuencias Alu. Estas secuencias solo se encuentran en los primates, por ello se han usado para diferenciar las células humanas de las murinas presentes en un modelo PDX. En esta figura, podemos observar que las células humanas (moradas) se encuentran formando estructuras ductales (con irregulares causadas por el proceso de tumorigénesis), mientras que las células no humanas (rosas), en nuestro caso células de ratón, constituyen el estroma.

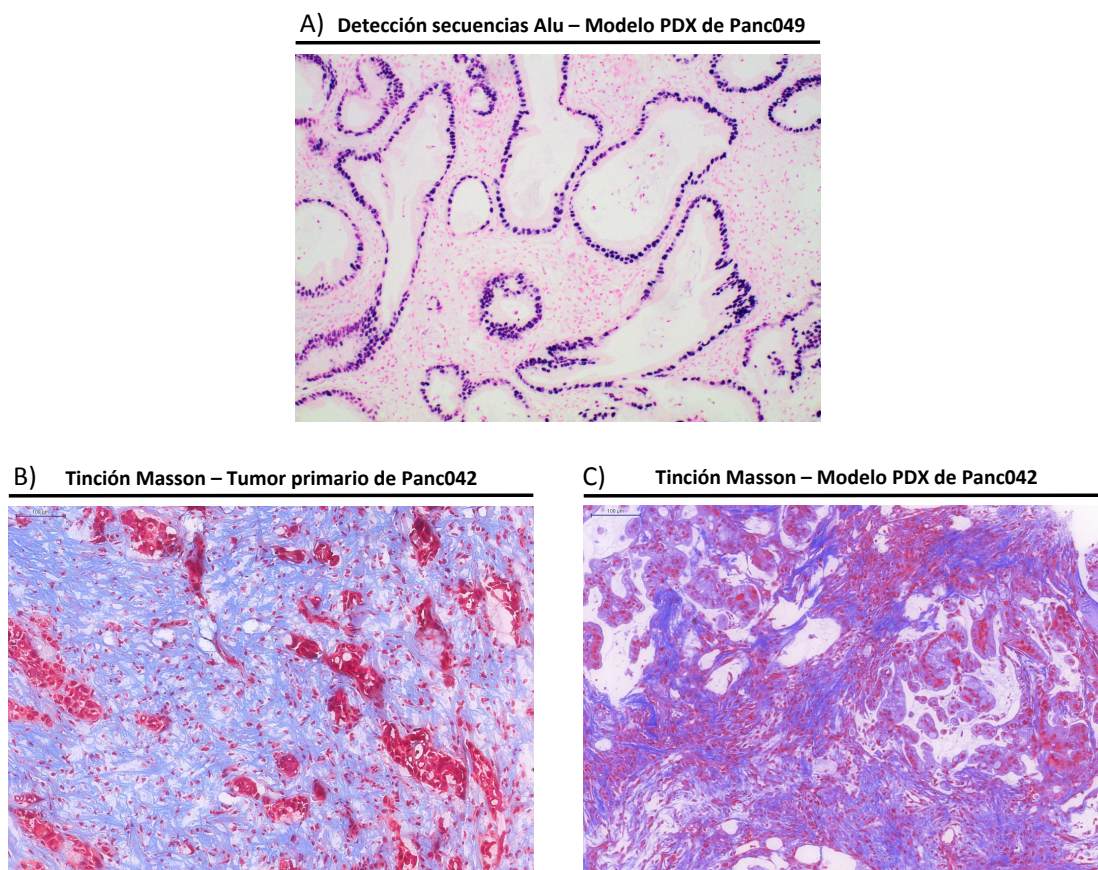


Figura 15. **A)** Detección de secuencias Alu (presentes solo en primates) en el modelo PDX de Panc049. Imagen cedida por el Dr. López-Casas (datos no publicados). Los tonos morados representan las células que presentan las secuencias Alu, es decir, las células humanas, mientras que los tonos rosas son células no humanas (en este caso son células murinas). **B)** Tinción tricrómica de Masson del tumor primario de Panc042. Presenta un 51.62% de tejido estromal. **C)** Tinción tricrómica de Masson del modelo PDX de Panc042. Presenta un 25.10% de tejido estromal. La muestra B estaba congelada mientras que la C se encontraba en un bloque de parafina, por ello ambas tinciones muestran tonalidades diferentes. Los tonos azules son las zonas de fibrosis, mientras que los tonos entre lila y rojo representan el resto de tejidos.

Adicionalmente, nuestras observaciones experimentales, indican que la cantidad de estroma sufre una disminución importante respecto a los tumores primarios humanos, en los que se sabe que se produce una gran reacción desmoplásica. En las Figuras 15B y C podemos ver muestras de un paciente (tumor primario y modelo PDX) sobre las

que se ha realizado una tinción tricrómica de Masson. Esta tinción permite diferenciar la zona fibrótica del resto de tejidos, es decir, que podemos cuantificar la cantidad de estroma de una muestra (ver más detalles en la sección Materiales y métodos). La cuantificación de la muestra de tumor primario indica que posee un 51.62% de estroma (humano), mientras que el modelo PDX generado a partir de ese tumor primario contiene solo un 25.10% de tejido estromal (murino).

Se considera que las muestras de tumor primario están siempre “contaminadas” por células no tumorales, tanto pertenecientes al estroma como al sistema inmune. En ese sentido, los perfiles de expresión de los modelos PDX, establecidos en plataformas de hibridación diseñadas para muestras humanas (que carecen de sondas específicas para genes de ratón), deberían tener una menor representación del componente del estroma de ratón que del componente tumoral humano.

Para cuantificar el número de sondas que puedan unirse a tejido murino de la plataforma de hibridación usada en esta tesis, utilizamos la herramienta *Xenome* [17]. Esta herramienta nos permite cuantificar el porcentaje de similitud de la secuencia de cada sonda, con las secuencias murinas que potencialmente podrían estar contaminando la muestra. Hemos analizado 1093500 secuencias (ver detalles en la sección Materiales y métodos), correspondientes a los oligonucleótidos de las 54675 sondas que hemos usado para este estudio. El programa *Xenome* ha detectado 19920 secuencias que pueden considerarse ambiguas, 19300 que se unirían indistintamente al genoma de humano y ratón, y 620 que se unen específicamente al genoma murino. Esas secuencias ambiguas constituyen un porcentaje mínimo del total de secuencias analizadas (1.82%). Por tanto, consideramos que la plataforma de hibridación (*GeneChip® Human Genome UI33 Plus 2.0 Array*), es suficientemente robusta para analizar los perfiles de expresión de los modelos PDX.

La especificidad de especie de las plataformas de hibridación y la reducción de los componentes estromales en modelos PDX, deberían ser dos factores suficientes para reducir drásticamente la detección de esos componentes no tumorales en dichos modelos. Adicionalmente, la ausencia de contaminación no tumoral en los modelos

PDX, podría desembocar en una sobreestimación de las diferencias entre estos modelos y los tumores primarios, donde la contaminación no tumoral humana es inevitable. Por ello, investigamos si este efecto de contaminación no tumoral podría ser suficiente para explicar la separación de las muestras en el segundo componente del espacio de expresión génica de tumores primarios y modelos PDX (eje Y de la Figura 10, el cual separaba las muestras en función del huésped).

Para dilucidar este punto, hemos usado el método ESTIMATE [127], el cual nos permite inferir la fracción estromal e inmune en las diferentes muestras. ESTIMATE está basado en una firma génica que contiene genes característicos expresados en las células humanas no tumorales infiltrantes (NTI), tanto del estroma como del sistema inmune. Según este método, los tumores primarios de ADP muestran una mayor proporción de células NTI que los tumores primarios de CHC (Figura 16).

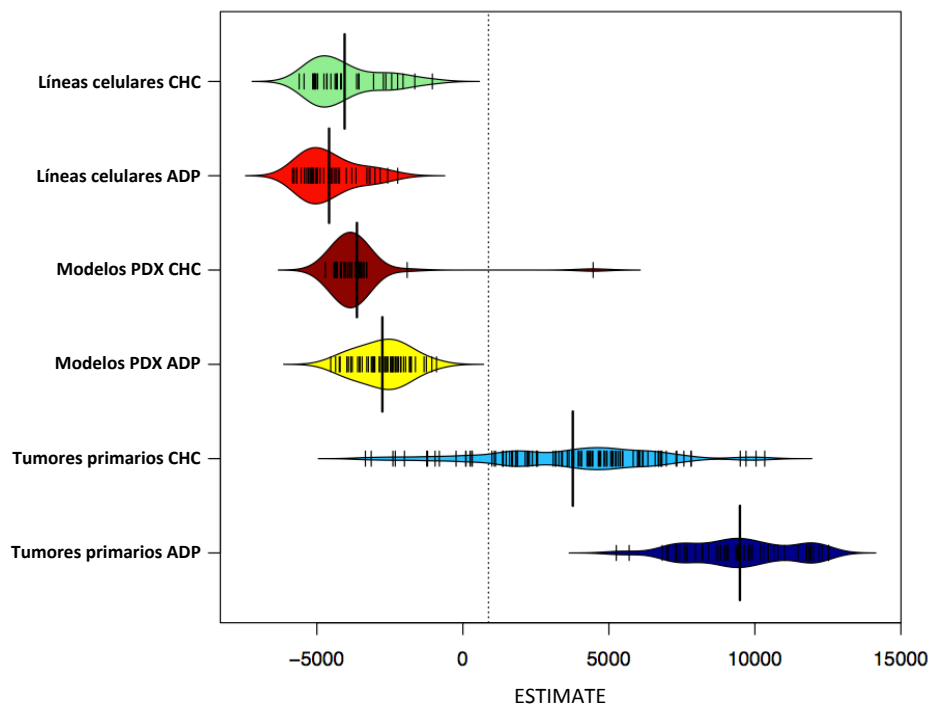


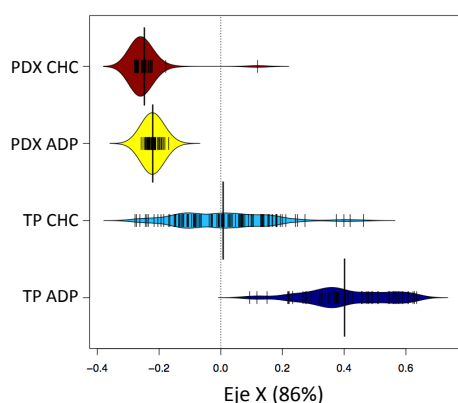
Figura 16. *Beanplot* de los valores ESTIMATE de los seis grupos de muestras usados en el estudio. Los valores altos indican una gran proporción de células NTI y viceversa. El área coloreada representa la densidad de muestras, las líneas negras verticales y pequeñas corresponden a cada muestra, la línea negra más ancha de cada

grupo representa el valor medio del mismo y la línea vertical punteada corresponde a la media global.

Adicionalmente, podemos observar que ambos tipos de modelos PDX y de líneas celulares, presentan valores de ESTIMATE muy bajos (negativos), un resultado coherente con lo esperado que indica que las células NTI están prácticamente ausentes en ambos modelos. Este dato puede interpretarse como una corroboración de nuestras observaciones experimentales realizadas en los modelos PDX respecto al estroma (Figura 15).

A continuación, hemos generado espacios de expresión usando las mismas muestras pero considerando solo las sondas de los 282 genes presentes en la firma de las células NTI. En los espacios obtenidos para ambas combinaciones (tumores primarios y modelos PDX o tumores primarios y líneas celulares) solo el primer eje es significativo. En ambos casos, el primer eje explica la mayor parte de la varianza, por lo que hemos decidido representar las coordenadas en un único eje mediante un *beanplot* (Figura 17).

A) Tumores primarios y modelos PDX según los genes presentes en la firma de las células NTI (Eje X)



B) Tumores primarios y líneas celulares según los genes presentes en la firma de las células NTI (Eje X)

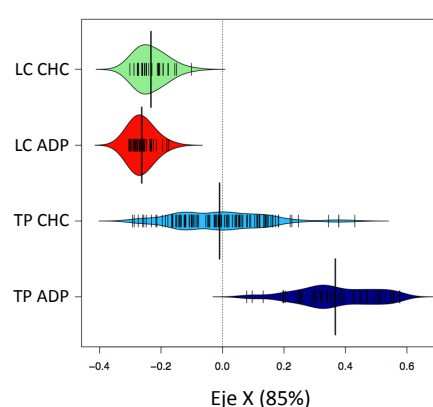
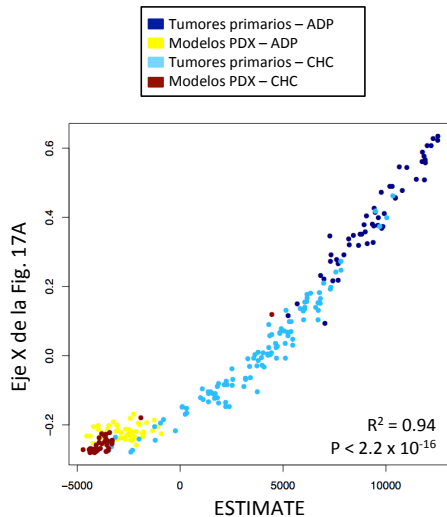


Figura 17. Contribución de las células NTI. **A)** Coordenadas del eje X del espacio de expresión génica de los tumores primarios (TP) y los modelos PDX (PDX), cuando solo se usan las sondas de los genes presentes en la firma de las células NTI (el

segundo eje del MCA no es estadísticamente informativo, $p = 0.06$). **B)** Coordenadas del eje X del espacio de expresión génica de los tumores primarios (TP) y las líneas celulares (LC), cuando solo se usan las sondas de los genes presentes en la firma de las células NTI (el segundo eje del MCA no es estadísticamente informativo, $p = 0.06$). En ambos casos, el área coloreada representa la densidad de muestras en el espacio unidimensional, las líneas negras verticales y pequeñas corresponden a cada muestra, la línea negra más ancha de cada grupo representa el valor medio del mismo y la línea vertical punteada corresponde a la media global.

A) ESTIMATE vs. Eje X del espacio de expresión génica de tumores primarios y modelos PDX (Fig. 17A)



B) ESTIMATE vs. Eje X del espacio de expresión génica de tumores primarios y líneas celulares (Fig. 17B)

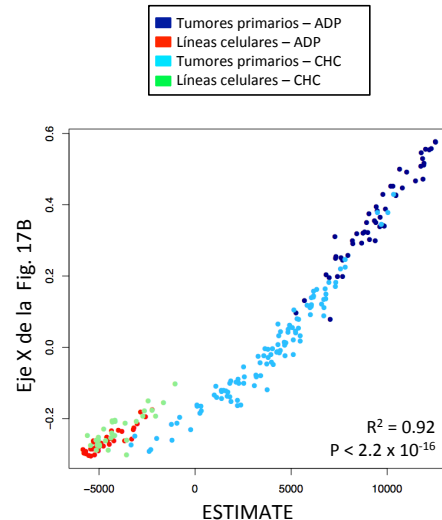


Figura 18. Gráficos que muestran la correlación entre los valores ESTIMATE y la contribución de las células NTI. **A)** Correlación entre el valor ESTIMATE y el eje X de la Figura 17A (MCA de tumores primarios y modelos PDX cuando solo usamos las sondas de los genes de la firma de las células NTI). **B)** Correlación entre el valor ESTIMATE y el eje X de la Figura 17B (MCA de tumores primarios y líneas celulares cuando solo usamos las sondas de los genes de la firma de las células NTI).

La separación del primer componente del MCA, en ambos casos, es muy similar a la separación realizada en función de los valores ESTIMATE. De hecho, si representamos la correlación del primer componente del espacio (eje X) respecto a los

valores ESTIMATE, en ambos casos, el valor de dicha correlación es muy significativo (Figura 18). Por tanto, estos resultados refuerzan la idea de que los perfiles de expresión de las muestras están influenciados por la cantidad de estroma detectado por la plataforma de hibridación. Es decir, en los tumores primarios la contaminación estromal está presente y es detectada, mientras que en los modelos PDX el estroma murino no se detecta (en las líneas celulares no hay estroma).

Para evaluar el efecto de las células NTI en nuestros resultados, hemos generado los espacios de expresión génica de tumores primarios y modelos PDX o líneas celulares (mismas muestras que en las Figuras 10 y 13, respectivamente), pero excluyendo del análisis las sondas de los genes de la firma de las células NTI (Figura 19).

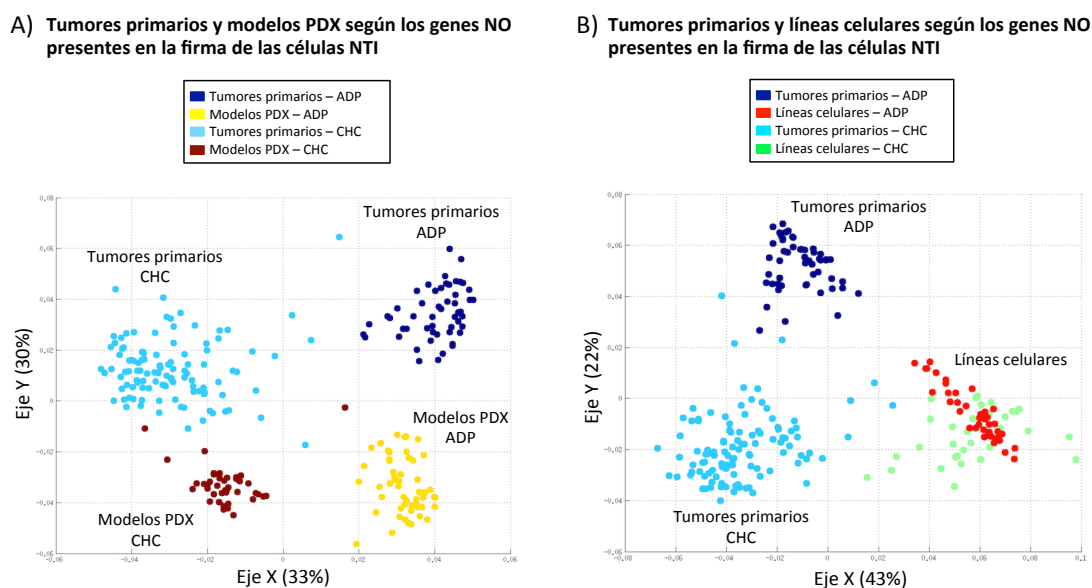


Figura 19. Espacios de expresión génica excluyendo las sondas de los genes presentes en la firma de las células NTI. **A)** Tumores primarios y modelos PDX (mismas muestras que la Figura 10). Test de Wilcoxon: entre los ejes X e Y, $p = 4.40 \times 10^{-22}$; entre los ejes Y y Z, $p = 6.90 \times 10^{-4}$. Solución óptima para $k = 4$ (ICH = 268.35). **B)** Tumores primarios y líneas celulares (mismas muestras que la Figura 13). Test de Wilcoxon: entre los ejes X e Y, $p = 3.99 \times 10^{-13}$; entre los ejes Y y Z, $p = 0.14$. Solución óptima para $k = 3$ (ICH = 373.85). En ambas figuras, cada punto representa el patrón de expresión de una muestra. El primer eje de separación es el eje X, el segundo el eje Y.

La eliminación de las sondas de esos 282 genes tiene un efecto mínimo en la distribución de las muestras del espacio de expresión resultante. Por tanto, esos genes no tienen una influencia fundamental en la separación de las muestras en el espacio, lo que indica que hay más genes que contribuyen a la separación de las muestras en el eje X del espacio generado con todos los genes (Figura 10).

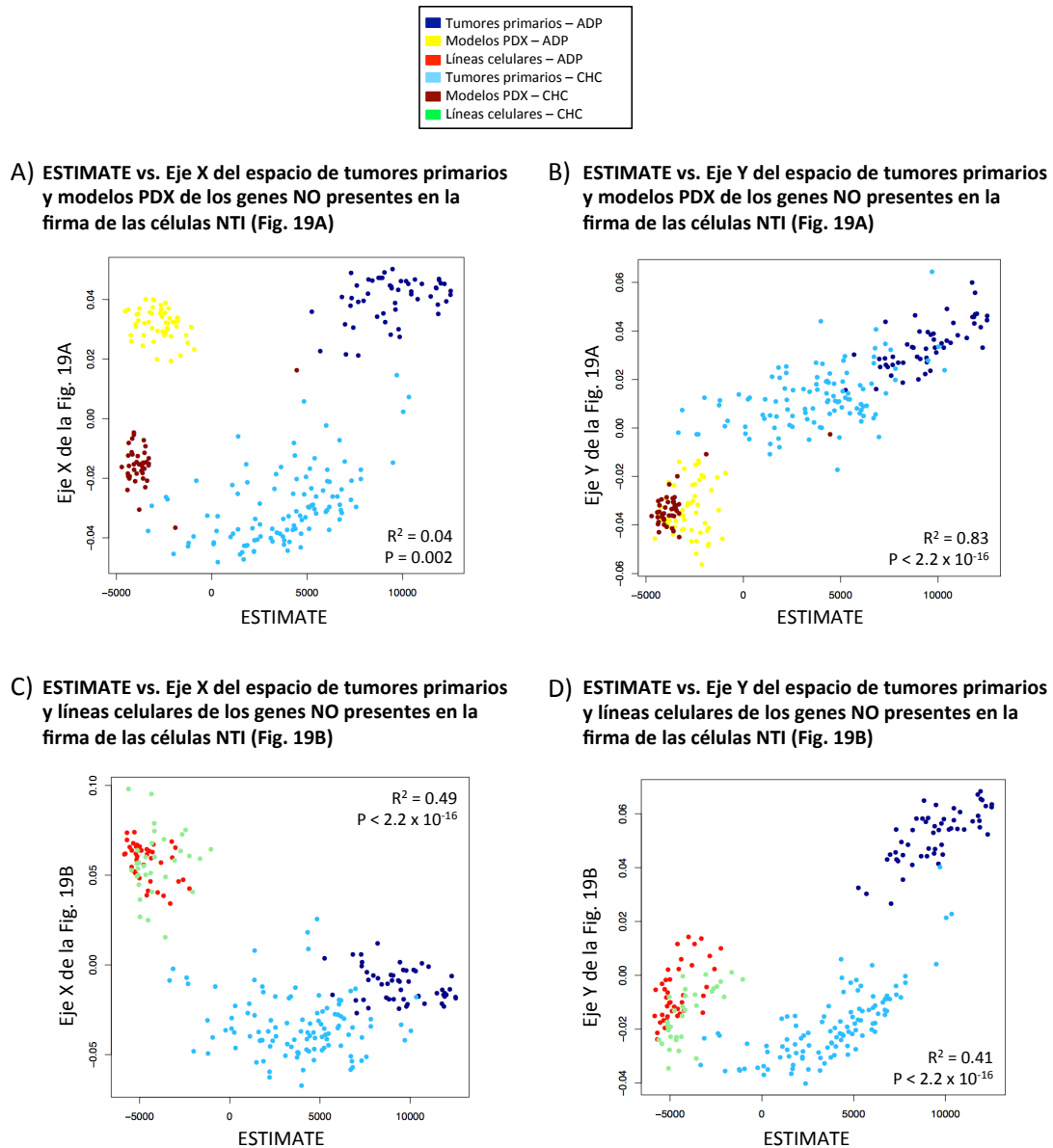


Figura 20. Correlación de los valores ESTIMATE respecto a los ejes X e Y de los espacios de expresión obtenidos excluyendo las sondas de los genes presentes en la firma de las células NTI. **A)** Valores ESTIMATE frente al eje X de la Figura 19A (MCA de tumores primarios y modelos PDX). **B)** Valores ESTIMATE frente al eje Y

de la Figura 19A (MCA de tumores primarios y modelos PDX). **C)** Valores ESTIMATE frente al eje X de la Figura 19B (MCA de tumores primarios y líneas celulares). **D)** Valores ESTIMATE frente al eje Y de la Figura 19B (MCA de tumores primarios y líneas celulares).

A continuación, comparamos los ejes X e Y de ambos espacios (tumores primarios y modelos PDX o tumores primarios y líneas celulares, Figura 19) con los correspondientes valores ESTIMATE (Figura 20). En la figura 20B, se puede observar una llamativa y significativa correlación entre el eje Y del espacio que incluye los modelos PDX, y los valores ESTIMATE ($R^2 = 0.83$, $p < 2.2 \times 10^{-16}$). Este resultado indica que las células NTI contribuyen a la separación de las muestras en función del huésped, es decir, a la separación de los tumores primarios respecto de los modelos PDX (separación observada previamente en los ejes Y de las Figuras 10 y 19A).

En resumen, los modelos PDX conservan parte de la expresión de sus tumores de origen, pero también se observan diferencias debidas, en parte, a la cantidad de células estromales y del sistema inmune presentes en las muestras humanas. En los modelos PDX, estas células infiltrantes corresponden al huésped murino y no se detectan con las plataformas de hibridación diseñadas para humanos (como ya hemos demostrado anteriormente).

Estos resultados nos llevan a proponer un experimento *in silico*, en el espacio de expresión de tumores primarios y modelos PDX, en el que intentamos corregir el efecto de las células NTI. Para ello, hemos utilizado un modelo de regresión lineal simple para cada componente del espacio, con los valores ESTIMATE como variable independiente (ver detalles en la sección Materiales y métodos). Tras la eliminación del efecto asociado a las células NTI, hemos generado un nuevo espacio de expresión con los residuos resultantes. Dicho espacio se puede considerar “corregido por el efecto de las células NTI” (Figura 21). En contraste con el espacio original (Figura 10), el espacio corregido no muestra una separación en el segundo eje, es decir, los

modelos PDX no se separan de los tumores primarios. En el caso de las muestras derivadas de páncreas, los tumores primarios y los modelos PDX son indistinguibles, al igual que en el caso de las muestras derivadas de hígado, donde ambos tipos de muestras solapan formando un único grupo.

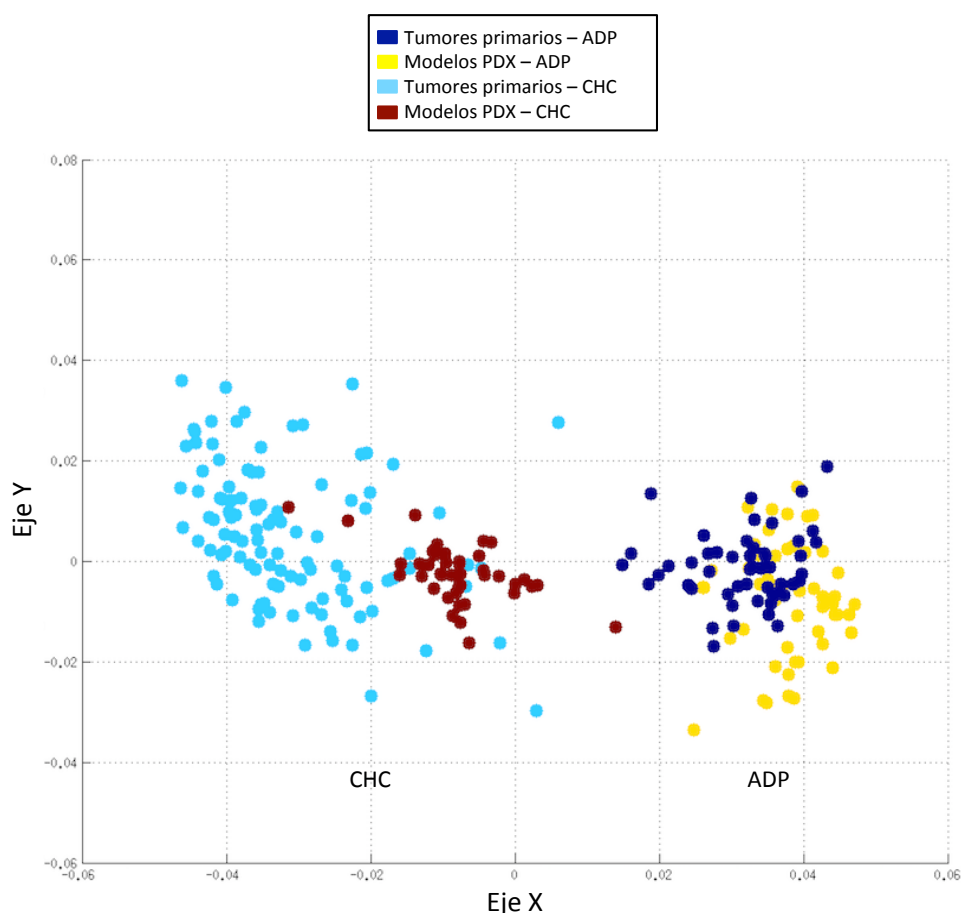


Figura 21. Espacio de expresión génica de tumores primarios y modelos PDX tras la corrección (*in silico*) del efecto de las células NTI. Las coordenadas de cada eje corresponden a los residuos obtenidos de la regresión lineal de cada componente del espacio de expresión génica de la Figura 19A (espacio obtenido excluyendo las sondas de los genes presentes en la firma de las células NTI).

Este resultado avala que la principal fuente de variación entre los perfiles de expresión de modelos PDX y tumores primarios, es la presencia de células NTI en las muestras humanas (tumores primarios) y su detección en la plataforma de hibridación.

4.1.5. Efecto de la microdissección en la contribución de las células no tumorales infiltrantes a los perfiles de expresión de los tumores primarios

Como hemos visto, las células NTI tienen un efecto importante en los perfiles de expresión de los tumores primarios. Además, hemos demostrado que las plataformas de hibridación son lo suficientemente específicas para no detectar el estroma murino en el caso de los modelos PDX. A pesar de ello, muchos investigadores proponen que la mejor solución es eliminar físicamente las células NTI, y cualquier otra contaminación de una muestra tumoral, mediante microdissección de las células tumorales. La microdissección elimina el componente no tumoral de la muestra, que normalmente tiene que ser fijada y procesada en secciones para poder visualizarse con un microscopio acoplado a un sistema de análisis de imagen (para seleccionar las células de interés) y un sistema de microdissección por láser.

Hemos querido comprobar como se comportan ese tipo muestras en un espacio de expresión génica respecto al resto de muestras, pues nos preguntamos si sus perfiles de expresión serían más similares a los tumores primarios completos o a los modelos PDX. Para ello, hemos usado un conjunto de datos generados por Collisson et al. en 2011 [16], en el cual se microdisseccionaron las células tumorales antes de generar los perfiles de expresión.

En la Figura 22, se puede observar el espacio de expresión resultante al usar modelos PDX, tumores primarios completos y tumores primarios microdisseccionados. Las dos dimensiones añadidas al espacio por considerarse informativas, explican el 79% de la varianza (54% el primer eje, eje X, y 25% el segundo, eje Y; $p = 6.39 \times 10^{-11}$). La tercera dimensión no es significativa ($p = 0.12$). Claramente se pueden observar tres grupos, los tres grupos obtenidos de forma no supervisada según el *k-means* ($k = 3$, $ICH = 251.28$), correspondientes a los tumores primarios microdisseccionados, tumores primarios completos y modelos PDX.

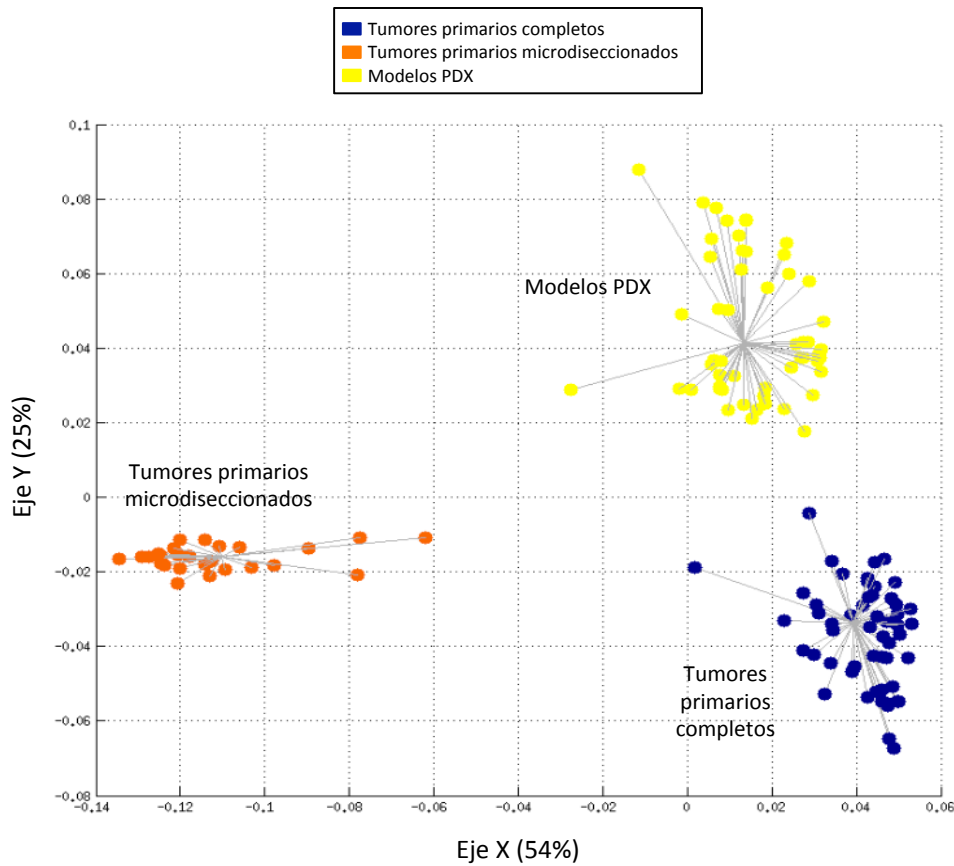


Figura 22. Espacio de expresión génica de los modelos PDX, los tumores primarios microdisseccionados y los tumores primarios completos. Cada punto representa el patrón de expresión de una muestra. El primer eje de separación es el eje X, el segundo es el eje Y. Solución óptima para $k = 3$ (ICH = 251.28).

Sorprendentemente, el eje que explica el mayor porcentaje de variabilidad (eje X, 54%) separa claramente los tumores microdisseccionados de los otros dos grupos. Este resultado sugiere que la microdissección de los tumores constituye una fuente de variabilidad adicional.

Cuando usamos el método ESTIMATE, para comparar la contribución de las células NTI en los tumores primarios completos, las muestras microdisseccionadas y los modelos PDX (Figura 23), observamos que la proporción de células NTI en las muestras microdisseccionadas es notablemente menor que en los tumores primarios completos, tal como era esperado. Los valores ESTIMATE son más altos que en los modelos PDX, lo cual se podría deber a una contaminación residual de células NTI.

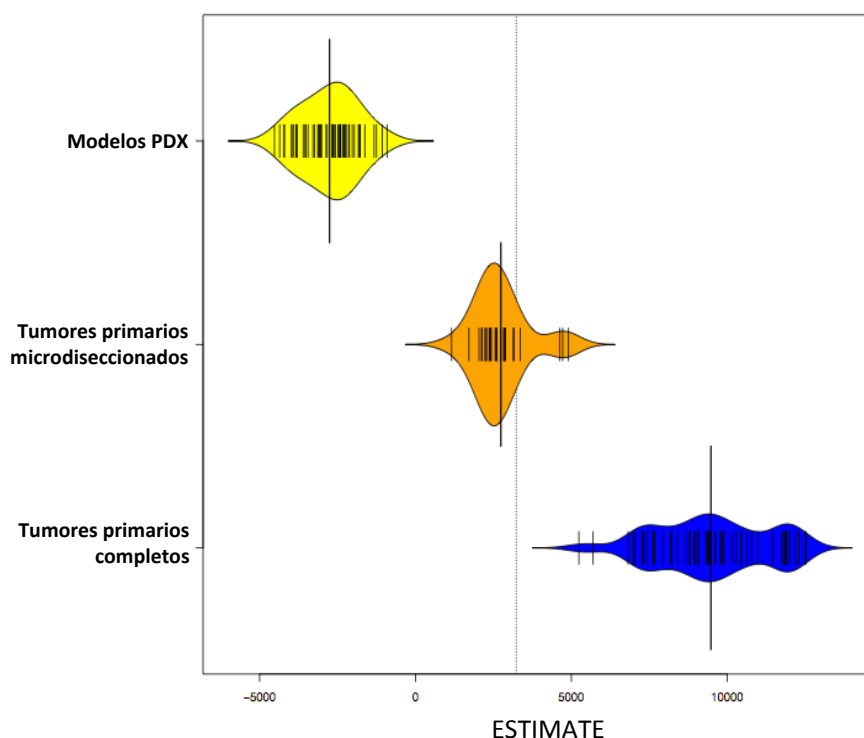


Figura 23. Valores ESTIMATE de los modelos PDX, los tumores primarios microdisseccionados y los tumores primarios completos. Los valores altos indican una gran proporción de células NTI y viceversa. El área coloreada representa la densidad de muestras, las líneas negras verticales y pequeñas corresponden a cada muestra, la línea negra más ancha de cada grupo representa el valor medio del mismo y la línea vertical punteada corresponde a la media global.

En la Figura 22, se observa que las muestras microdisseccionadas se encuentran a una mayor distancia de los tumores primarios completos que los modelos PDX. Además, el resultado mostrado en la Figura 23 sugiere que las diferencias entre muestras se podrían deber en parte al distinto porcentaje de células NTI. Para evaluar el efecto de las células NTI en estos resultados, hemos generado un nuevo espacio de expresión génica pero excluyendo del análisis las sondas de los genes de la firma de las células NTI (Figura 24).

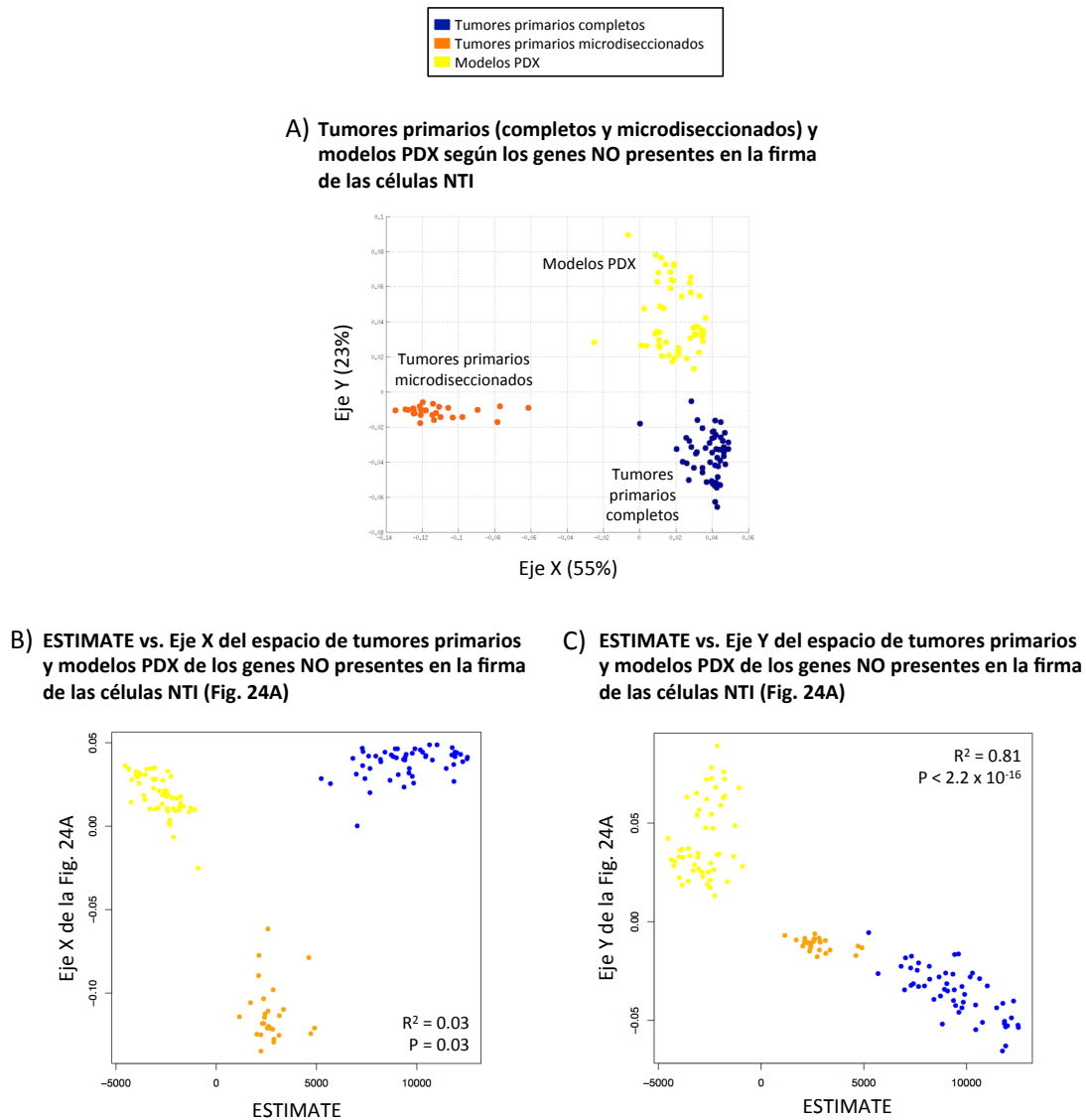


Figura 24. A) Espacios de expresión génica de tumores primarios (completos y microdisseccionados) y modelos PDX excluyendo las sondas de los genes presentes en la firma de las células NTI (mismas muestras que la Figura 22). Test de Wilcoxon: entre los ejes X e Y, $p = 3.91 \times 10^{-11}$; entre los ejes Y y Z, $p = 0.11$. Solución óptima $k = 3$ (ICH = 243.10). **B)** Correlación de los valores ESTIMATE respecto al eje X del espacio de expresión obtenido excluyendo las sondas de los genes presentes en la firma de las células NTI (Figura 24A). **C)** Correlación de los valores ESTIMATE respecto al eje Y del espacio de expresión obtenido excluyendo las sondas de los genes presentes en la firma de las células NTI (Figura 24A).

La eliminación de las sondas de los 282 genes de la firma de las células NTI, tiene un efecto prácticamente indistinguible (comparar Figura 24A y Figura 22), confirmando que el estroma no es la principal fuente de separación de los perfiles de expresión en el espacio. A continuación, examinamos la correlación entre los ejes de dicho espacio con los valores ESTIMATE (Figuras 24B-C). El eje X muestra una débil correlación con el valor de ESTIMATE ($R^2 = 0.03$, $p = 0.03$, Figura 24B), por lo que el efecto rescatado por este eje parece no estar influenciado por las células NTI. Este primer componente (eje X) muestra la separación de los tumores primarios microdisecccionados del resto de muestras. Esto indica que el procesamiento de la muestra, necesario para llevar a cabo la técnica de microdissección, está provocando cambios en el ARN, lo que modifica la detección de la expresión en la plataforma de hibridación.

Sin embargo, observamos que la correlación entre el eje Y del espacio y los valores ESTIMATE, es muy significativa ($R^2 = 0.81$, $p < 2.2 \times 10^{-16}$, Figura 24C). Este resultado sugiere que las células NTI contribuyen a la separación de las muestras en el eje Y, pero que no representan la mayor fuente de variabilidad entre las mismas.

Estos resultados nos llevan a proponer de nuevo el experimento de corrección *in silico*. En este caso, la corrección es aplicada sobre el espacio de expresión de tumores primarios, completos y microdisecccionados, y modelos PDX, en el cual hemos eliminado las sondas de los genes presentes en la firma de las células NTI (ver detalles en la sección Materiales y métodos). El espacio de expresión “corregido por el efecto de las células NTI” (Figura 25), muestra que al eliminar dicho efecto los tumores primarios completos y los modelos PDX solapan en un único grupo, separado de los tumores microdisecccionados.

Este resultado refuerza nuestra hipótesis de que la microdissección altera de algún modo el ARN de estas muestras y, en consecuencia, se altera la detección de los niveles de expresión en la plataforma de hibridación (pese a que los datos se normalicen por métodos robustos como los que hemos utilizado en esta tesis).

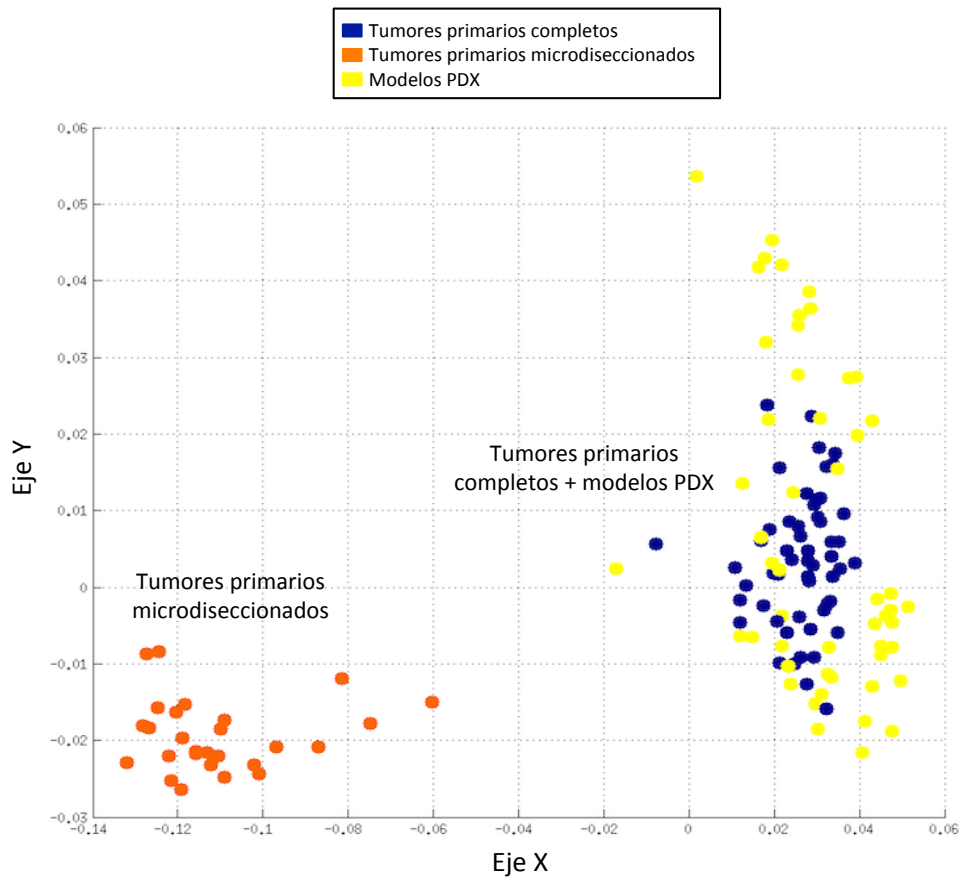


Figura 25. Espacio de expresión génica de tumores primarios (microdisseccionados y completos) y modelos PDX tras la corrección (*in silico*) del efecto de las células NTI. Las coordenadas de cada eje corresponden a los residuos obtenidos de la regresión lineal de cada componente del espacio de expresión génica de la Figura 24A (espacio obtenido excluyendo las sondas de los genes presentes en la firma de las células NTI).

Los perfiles de expresión de las muestras de tumores primarios microdisseccionados, fueron extraídos de un trabajo en el cual se usaban para establecer subtipos de cáncer de páncreas [16]. En ese trabajo se establecieron 3 subtipos: *classical*, *exocrine-like* y *quasimesenchymal*. Además, se definió una firma génica, denominada PDAssigner y constituida por 62 genes, capaz de identificar los subtipos en nuevas muestras, a partir de sus perfiles de expresión.

Las muestras microdisseccionadas fueron generadas como parte de dicho trabajo, pero los subtipos se establecieron a partir de un conjunto de datos formado por los perfiles

de expresión de los tumores primarios microdisseccionados y de tumores primarios completos (GSE15471, muestras que también forman parte de nuestro conjunto de datos de tumores primarios de ADP [3]). Dada la diferencia que presentan ambos tipos de muestras en los espacios de expresión génica, nos planteamos si los subtipos obtenidos pueden estar sesgados por las diferencias técnicas y la cantidad de células NTI que hemos observado. Para comprobar esta hipótesis, coloreamos el espacio de expresión obtenido con los tumores primarios completos, los tumores primarios microdisseccionados y los modelos PDX (Figura 22) en función de los subtipos proporcionados por los autores (Figura 26A).

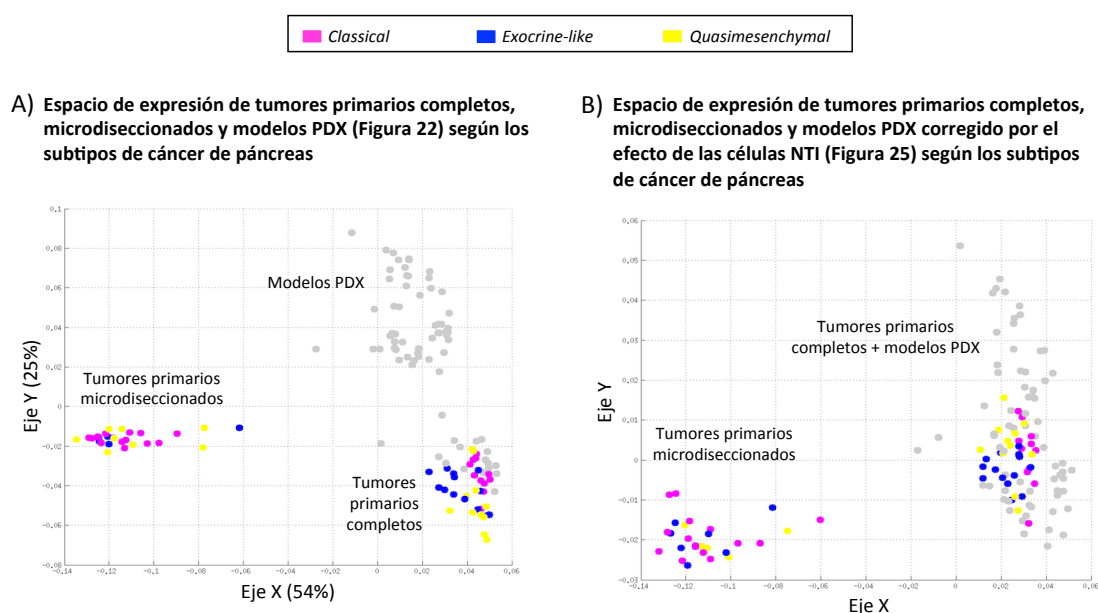


Figura 26. **A)** Espacio de expresión génica de los modelos PDX, los tumores primarios (microdisseccionados y completos) y modelos PDX (Figura 22), coloreado según la clasificación en subtipos de Collisson et al. Cada punto representa el patrón de expresión de una muestra. El primer eje de separación es el eje X, el segundo es el eje Y. **B)** Espacio de expresión génica de tumores primarios (microdisseccionados y completos) y modelos PDX tras la corrección (*in silico*) del efecto de las células NTI (Figura 25) y coloreado según la clasificación en subtipos de Collisson et al. Las coordenadas de cada eje corresponden a los residuos obtenidos de la regresión lineal de cada componente del espacio de expresión génica de la Figura 24A (espacio obtenido excluyendo las sondas de los genes presentes en la firma de las células NTI).

La Figura 26A no muestra un sesgo en los subtipos, por lo que parece que el sistema de normalización y eliminación de efectos de lote o *batch* aplicado por los autores ha sido eficaz. Sin embargo, nos resulta sorprendente que hayan podido definir dichos subtipos, cuando nosotros no somos capaces de detectar ningún tipo de agrupación en las muestras pertenecientes a los tres subtipos.

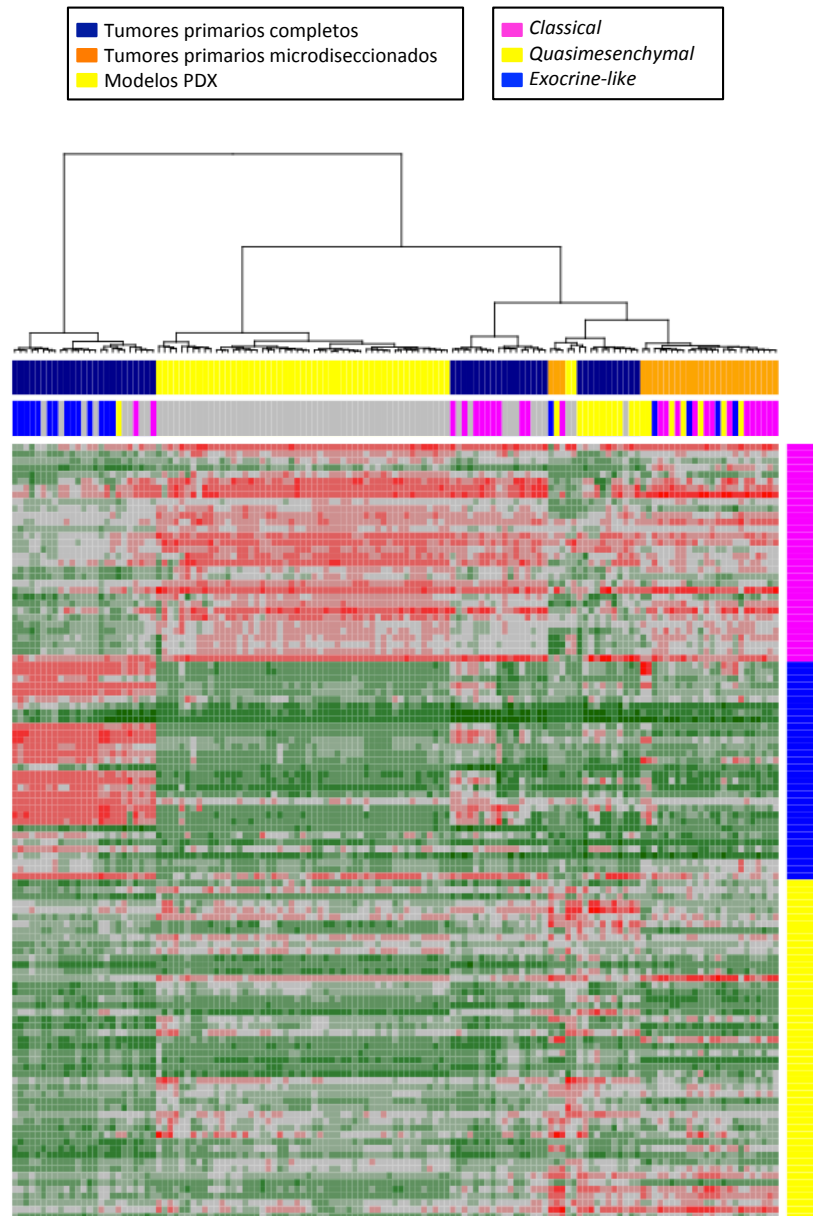
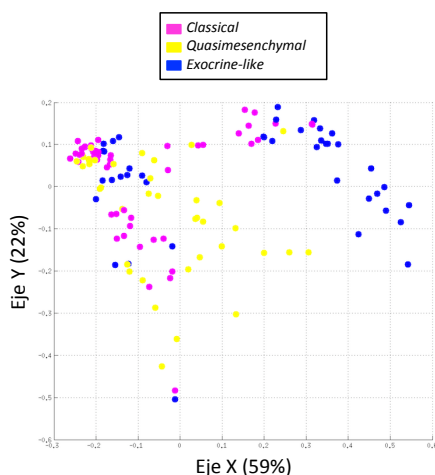


Figura 27. *Heatmap* de los valores de intensidad normalizados de las sondas pertenecientes a los genes de la firma PDAssigner. Las columnas son las muestras y las filas las sondas. Código de color del *heatmap*: rojo, intensidad alta; gris, intensidad media; verde, intensidad baja.

Por ello, decidimos examinar el patrón de expresión de los genes de la firma génica (PDAssigner) obtenida a partir de los subtipos, la cual debería reflejar mejor la agrupación de las muestras en los diferentes subtipos (Figura 27). La Figura 27, tampoco muestra una clasificación clara de los subtipos en función de los niveles de expresión de las sondas pertenecientes a los genes de la firma definida por los autores.

A continuación, estudiamos como se distribuyen las muestras en un espacio de expresión génica si solo tenemos en cuenta las sondas de la firma PDAssigner. El resultado se puede observar en la Figura 28A, donde las muestras del mismo subtipo siguen sin mostrar un patrón de agrupamiento, a pesar de estar usando solo los genes de la firma PDAssigner.

A) Espacio de expresión de tumores primarios completos, microdisseccionados y modelos PDX usando los genes de la firma PDAssigner y según los subtipos de cáncer de páncreas



B) Espacio de expresión de tumores primarios completos, microdisseccionados y modelos PDX usando los genes de la firma PDAssigner

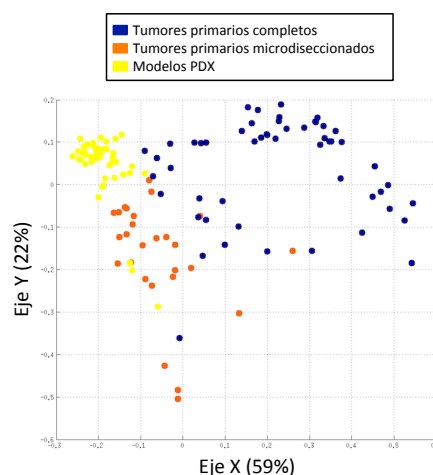


Figura 28. Espacio de expresión génica de los modelos PDX, los tumores primarios microdisseccionados y completos y modelos PDX (Figura 22). Cada punto representa el patrón de expresión de una muestra. El primer eje de separación es el eje X, el segundo es el eje Y. **A)** Figura coloreada según la clasificación en subtipos de Collisson et al. **B)** Figura coloreada en función del tipo de muestra.

La Figura 28B muestra el espacio de expresión coloreado según el tipo de muestra. Este espacio demuestra que al usar solo las sondas de la firma génica, las muestras se mezclan parcialmente en el espacio, en comparación con la Figura 22, pero todavía se pueden distinguir los tres tipos de muestras. Estos resultados implican que los subtipos de cáncer de páncreas identificados por los autores no son reproducibles cuando usamos su firma génica en nuestras muestras.

4.1.6. Análisis funcional de los modelos PDX de ADP

Hemos podido demostrar que los modelos PDX son más similares a los tumores primarios humanos a nivel transcripcional de lo que se creía. Sin embargo, nuestros análisis también corroboran que existen ciertas diferencias. Nuestros resultados indican que gran parte de estas diferencias, entre tumores primarios y modelos PDX, son debidas al efecto de las células NTI en las muestras humanas. A continuación, utilizamos la información de expresión para caracterizar en más detalle esas diferencias, centrándonos en la información de las rutas biológicas. Para ello, hemos usado los datos de expresión génica, previamente convertidos en valores binarios, y hemos calculado la diferencia de frecuencia de expresión de cada gen en los tumores primarios y en los modelos PDX de ADP. Es decir, hemos cuantificado el número de muestras en las que cada gen está expresado dentro de cada grupo (explicación más detallada en la sección Materiales y métodos). La Figura 29 muestra las diferencias de frecuencia de expresión obtenidas para cada gen.

La diferencia de frecuencia de expresión se utiliza para investigar que rutas biológicas pueden estar significativamente enriquecidas en los genes expresados diferencialmente y, por tanto, que procesos biológicos pueden asociarse a las diferencias entre modelos PDX y tumores primarios.

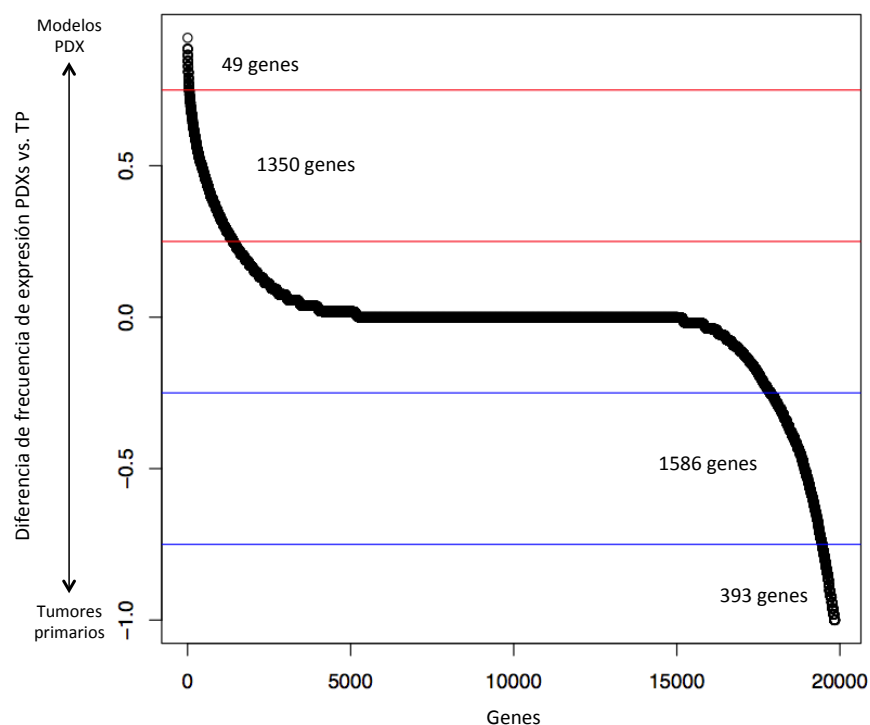


Figura 29. Diferencia de frecuencia de expresión entre modelos PDX (PDXs) y tumores primarios (TP). Cada punto representa un gen. Las líneas representan los valores 0.25 y 0.75 en rojo y -0.25 y -0.75 en azul. En cada tramo se representa el número de genes que contiene. Esta lista ordenada de genes fue utilizada para el análisis de enriquecimiento con GSEA.

El análisis de enriquecimiento (realizado con GSEA [111]) nos permitió identificar setenta y dos rutas, de la base de datos Reactome [78], que presentan acumulaciones significativas ($FDR < 0.05$) de genes sobre-expresados (aumentan su expresión) o reprimidos (genes que disminuyen sus niveles de expresión) en los modelos PDX (Figura 30). En concreto, 46 rutas biológicas presentaban una mayor presencia de genes sobre-expresados en modelos PDX y 26 rutas contenían principalmente una acumulación de genes reprimidos en dichos modelos. Estas rutas están relacionadas con un grupo más reducido de funciones celulares, lo que nos ha permitido realizar una agrupación de las mismas en grupos funcionales, siguiendo la jerarquía interna de la base de datos Reactome.

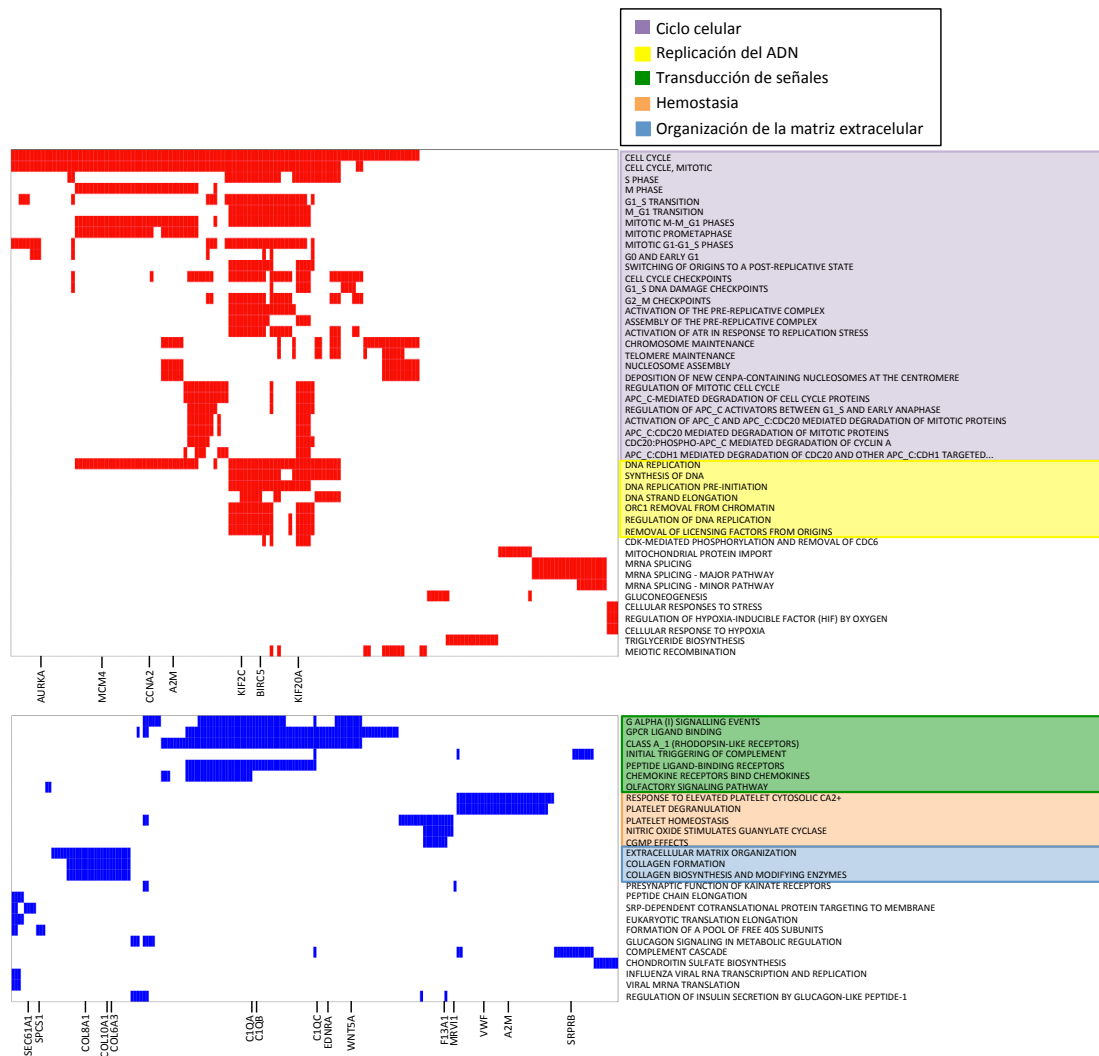


Figura 30. Grupos funcionales y rutas biológicas asociadas a los modelos PDX y la progresión tumoral. Hay 46 rutas sobre-expresadas (rojo) implicadas en 2 grupos funcionales y 26 rutas reprimidas (azul) pertenecientes a 3 grupos funcionales. Las áreas coloreadas representan las rutas implicadas en cada grupo funcional. Solo se muestran los genes más significativamente enriquecidos. La Tabla S2 (Anexo I – Material suplementario) contiene todos los genes significativamente enriquecidos en cada ruta, con los valores de frecuencia de expresión en los modelos PDX y en los tumores primarios.

Es particularmente interesante que los genes de las rutas relacionadas con ciclo celular y replicación del ADN están sobre-expresados en los modelos PDX, reflejando el aumento de proliferación en estos modelos, mientras que los genes de las rutas

relacionadas con transducción de señales, hemostasia y organización de la matriz extracelular están significativamente reprimidos, indicando que existe una adaptación al nuevo medio (Figura 30 y Tabla S2, Anexo I – Material suplementario). Parte de esa disminución en los niveles de expresión, también se debe a la sustitución del estroma humano por estroma murino. De hecho, hemos observado que los genes estromales, pertenecientes a la firma génica de células NTI, están significativamente enriquecidos en transducción de señales ($p = 1.6 \times 10^{-7}$), hemostasia ($p = 3.1 \times 10^{-5}$) y organización de la matriz extracelular ($p < 10^{-16}$). A su vez los genes del sistema inmune de dicha firma, se encuentran enriquecidos en las rutas implicadas en transducción de señales ($p = 6.5 \times 10^{-3}$) y hemostasia ($p = 0.02$). Esto implica que los cambios funcionales detectados están afectados por la contribución de las células NTI.

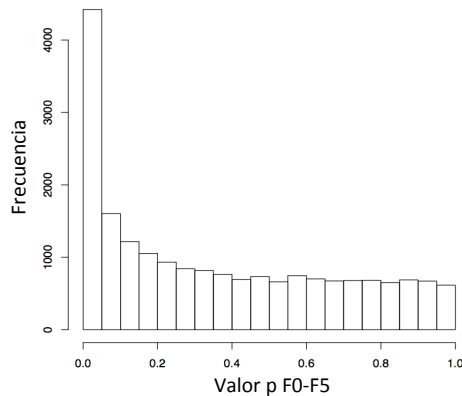
4.1.7. Estabilidad de los perfiles de expresión de los modelos PDX

Los modelos de PDX se mantienen con sucesivas rondas de implantación en los ratones receptores (se denomina “pase” a cada una de las veces que un tumor es reimplantado). Debido a que es posible que cada pase introduzca nuevos cambios de expresión, hemos examinado si los cambios funcionales previamente descritos se relacionan con el número de pases. Para ello, hemos utilizado un conjunto de datos adicional, que consta de perfiles de expresión de muestras de tumor primario (F0), modelos PDX de pase 5 (F5) y modelos PDX de pase 10 (F10) de 4 pacientes diferentes [30]. Debido a que este conjunto de datos es muy pequeño, no hemos usado la frecuencia de expresión, sino que hemos llevado a cabo un t-test utilizando el paquete *limma* de R (ver más detalles en la sección Materiales y métodos) para obtener los genes diferencialmente expresados.

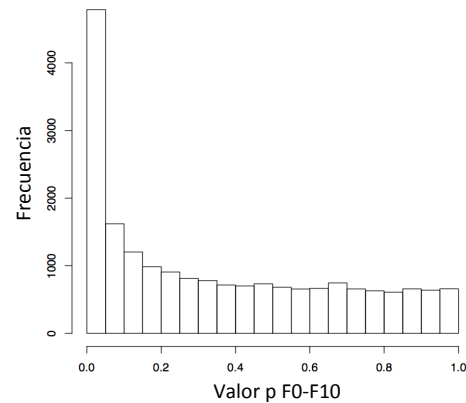
En el caso de las comparaciones F0 vs. F5 y F0 vs. F10, encontramos 17 y 19 genes diferencialmente expresados ($FDR < 0.05$), respectivamente, mientras que en el análisis F5 vs. F10 no obtuvimos ningún gen. El número tan bajo de genes

diferencialmente expresados puede explicarse por el hecho de que solo disponemos de tres muestras para hacer los análisis.

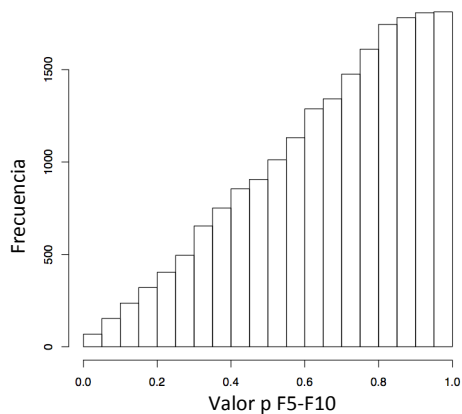
A) Distribución del valor p del análisis t-test F0 vs. F5



B) Distribución del valor p del análisis t-test F0 vs. F10



C) Distribución del valor p del análisis t-test F5 vs. F10



D) Representación del valor t del análisis t-test F0 vs. F5 frente al valor t de F0 vs. F10

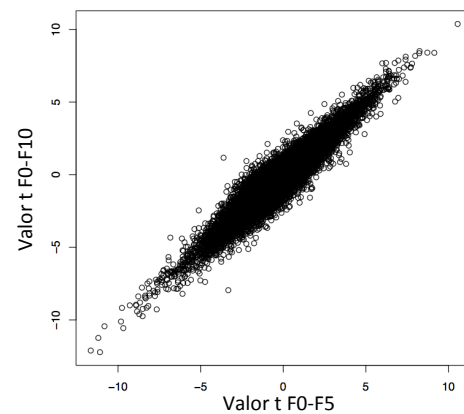


Figura 31. Representación de los resultados de los análisis t-test, mediante *limma*, entre F0, F5 y F10. **A)** Histograma de la distribución del valor p del análisis F0 vs. F5. **B)** Histograma de la distribución del valor p del análisis F0 vs. F10. **C)** Histograma de la distribución del valor p del análisis F5 vs. F10. **D)** Gráfico del valor t del análisis F0 vs. F5 frente al valor t de la comparación F0 vs. F10. Cada punto representa un gen.

Si observamos la distribución del valor p de las comparaciones F0 vs. F5 y F0 vs. F10 (Figuras 31A y 31B, respectivamente), podemos observar que la mayoría de los valores se encuentran cercanos a la significancia ($p < 0.05$). Sin embargo, la distribución del valor p de la comparación F5 vs. F10 (Figura 31C) indica que la mayoría de los valores se encuentran cercanos a 1. Cuando comparamos los valores del estadístico t de los análisis F0 vs. F5 y F0 vs. F10 (Figura 31D), podemos observar una correlación muy alta.

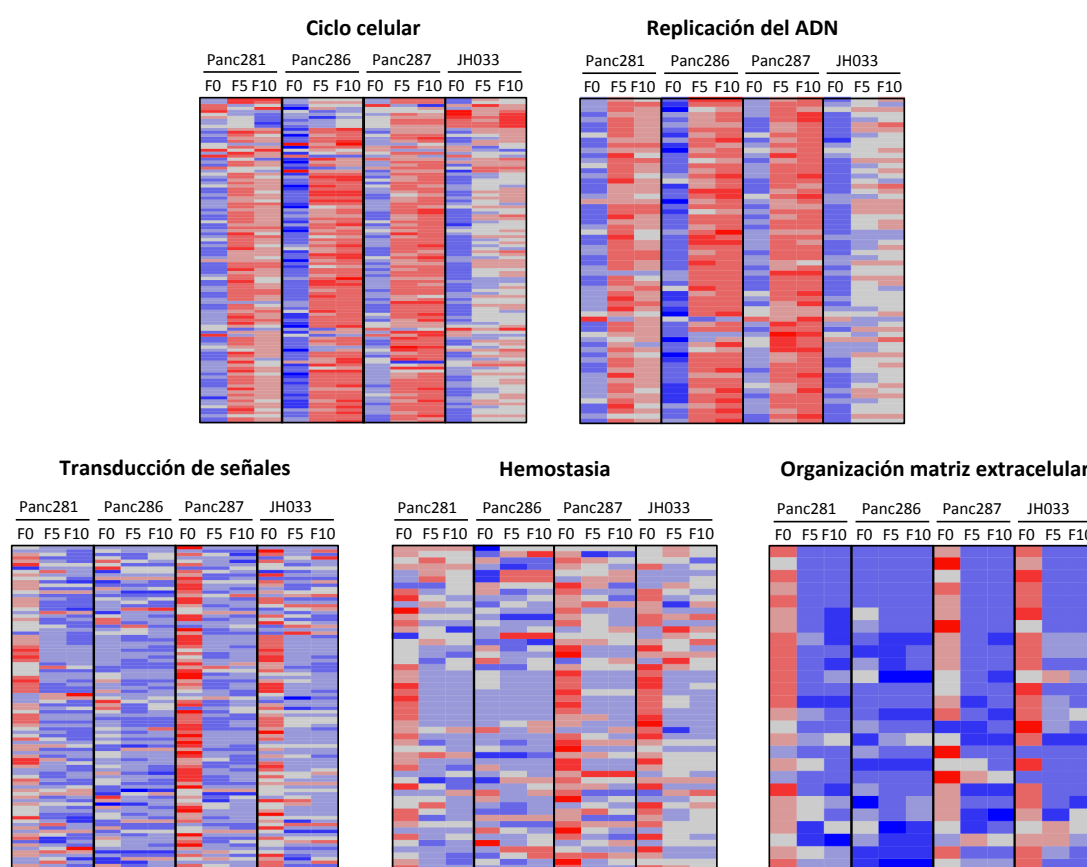


Figura 32. Resultados comparando tumores primarios y modelos PDX a diferentes pases. *Heatmaps* de los valores de expresión estandarizados (a nivel de gen) de los genes pertenecientes a las rutas alteradas en los modelos PDX y agrupados según grupos funcionales. Cada gráfico representa los genes significativamente enriquecidos en las rutas de cada grupo funcional. F0 (pase 0, tumor primario original), F5 (pase 5) y F10 (pase 10). Código de colores: rojo, alto; gris, medio; azul, bajo.

Estos datos ponen de manifiesto que se producen diferencias al generar el modelo PDX, pero que posteriormente no se producen cambios importantes, al menos entre el pase 5 y el 10.

En la sección anterior, hemos obtenido varios grupos funcionales (ciclo celular, replicación del ADN, transducción de señales, hemostasia y organización de la matriz extracelular) alterados en los modelos PDX respecto a los tumores primarios. A continuación, nos centramos en los genes de esos grupos funcionales (debido a muestran un cambio entre ambos modelos) y estudiamos si las diferencias observadas en los modelos PDX muestran cambios significativos entre los pases F5 y F10. La Figura 32 muestra los niveles de expresión estandarizados (a nivel de gen) de los genes significativamente enriquecidos en las rutas de cada grupo funcional.

Usando la herramienta GSEA, hemos confirmado que todos los grupos funcionales están alterados significativamente después de la implantación en los modelos PDX (F0 vs. F5 y F0 vs. F10). Sin embargo, ninguno de los grupos funcionales muestra cambios significativos entre los pases disponibles, F5 y F10 (Figura 33).

En resumen, aunque hemos demostrado que los perfiles de expresión de los modelos PDX mantienen similitudes con los perfiles de expresión de los tumores primarios, también hemos podido detectar ciertas diferencias causadas por la adaptación al nuevo ambiente murino. Las principales diferencias a nivel funcional se encuentran en los procesos de señalización y proliferación. Además, nuestros resultados sugieren que dichas alteraciones funcionales son estables a lo largo de los pases, lo cual avala el modelo PDX como modelo preclínico y de investigación.

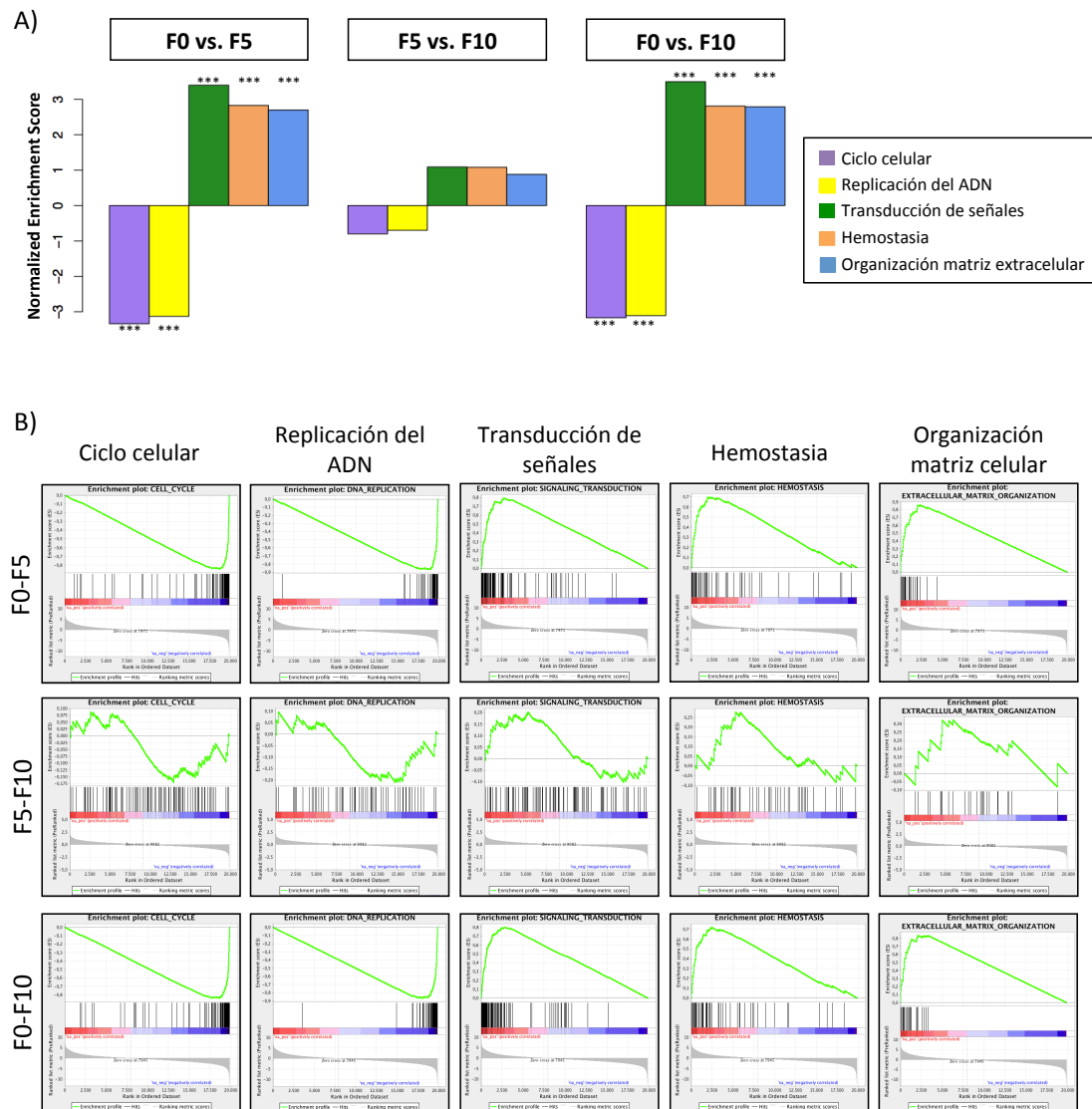


Figura 33. Resultados comparando tumores primarios y modelos PDX a diferentes pases. **A)** Los gráficos de barras muestran los valores NES (*Normalized Enrichment Score*) proporcionados por la herramienta GSEA en el análisis de los grupos funcionales. Representamos F0 vs. F5 (izquierda), F5 vs. F10 (medio) y F0 vs. F10 (derecha). El código de colores es el mismo que se ha usado para los grupos funcionales en la Figura 30. Los grupos funcionales que están significativamente enriquecidos se muestran con *** (FDR < 0.005). **B)** Gráficos de enriquecimiento de los grupos funcionales. Las filas representan los diferentes análisis entre pases. Las columnas son los diferentes grupos funcionales. Las comparaciones F0 vs. F5 y F0 vs. F10 son significativas (FDR < 0.05) para todos los grupos funcionales, mientras que F5 vs. F10 no es significativa para ninguno de ellos.

4.2. Modelos PDX derivados de metástasis

4.2.1. Análisis de los grupos funcionales en metástasis humanas

Muchas de las rutas biológicas alteradas en los modelos PDX son rutas típicamente desreguladas en tumores metastáticos respecto a tumores primarios, como por ejemplo ciclo celular y organización de la matriz extracelular. El grupo del Dr. Manuel Hidalgo propuso en 2011 [30], que los modelos PDX mimetizan la agresividad o determinadas características de los tumores metastáticos derivados de los tumores primarios, una idea que había sido sugerida previamente por Stratford et al. en 2010 [109] y que ha sido estudiada posteriormente por otros autores [88]. Por tanto, nuestra hipótesis es que el tumor injertado podría encontrarse en una situación de “metástasis forzada”, dado que necesita colonizar el nuevo ambiente para sobrevivir y además adquiere un estado muy elevado de proliferación. Por ello, nos planteamos si existe una similitud a nivel funcional entre metástasis y modelos PDX.

Para analizar esta similitud utilizamos los grupos funcionales. Primero, hemos obtenido de las bases de datos públicas conjuntos de datos que incluyen tumores primarios y metástasis. Los datos disponibles no son muy abundantes y proceden de metástasis en distintos órganos, por lo que hemos creado grupos en función del tejido/órgano de colonización. Se han realizado análisis de enriquecimiento (GSEA) de los grupos funcionales alterados en los modelos PDX (ver sección 4.1.6), comparando los genes diferencialmente expresados en cada grupo de metástasis respecto a los correspondientes tumores primarios.

Los resultados de estos análisis se resumen en la Tabla 5. Aunque se observan algunas diferencias, en general, las rutas biológicas alteradas en los modelos PDX también se encuentran significativamente alteradas en la misma dirección en las metástasis.

	Ciclo celular	Replicación del ADN	Transducción de señales	Hemostasia	Organización de la matriz extracelular
Met vs. TP en grasa [12]	0.007	0.001	0.280	0.079	0.063
Met vs. TP en hígado [12]	0.993	0.654	0.410	0.014	0.874
Met vs. TP en pulmón [12]	0.999	0.955	0.013	0.477	0
Met vs. TP en nodo linfático [12]	0.018	0.036	0.233	0.994	0.121
Met vs. TP en músculo [12]	0.016	0.016	0.004	0.112	0.054
Met vs. TP en hígado [117]	0.959	0.862	0	0.074	0
Met vs. TP en peritoneo [117]	0.037	0.046	0	0.001	0.001

Tabla 5. Valores FDR resultantes de los análisis de enriquecimiento entre las muestras de metástasis (Met) vs. tumor primario (TP), llevados a cabo en cada grupo funcional. Código de colores: morado oscuro, rutas sobre-expresadas y FDR < 0.05; morado claro, rutas sobre-expresadas y FDR < 0.25; azul oscuro, rutas infra-expresadas y FDR < 0.05; azul claro, rutas infra-expresadas y FDR < 0.25.

Este resultado sugiere que la hipótesis planteada es cierta, es decir, que los modelos PDX se encuentran en una situación de metástasis forzada y, por ello, las alteraciones funcionales son las mismas.

4.2.2. Espacios de expresión génica de modelos PDX derivados de tumor y de metástasis

Nuestros resultados muestran que las rutas alteradas en los modelos PDX también están significativamente alteradas en metástasis humanas. A continuación, nos centramos en si también existen similitudes entre los modelos PDX derivados de tumor primario y modelos PDX derivados de metástasis. Para ello, se generaron

nuevos perfiles de expresión de metástasis frescas humanas y modelos PDX derivados de metástasis. A dichas muestras, le añadimos los perfiles de expresión de los tumores primarios y los modelos PDX derivados de tumor primario generados para este estudio. Adicionalmente, combinamos las muestras con el conjunto de datos público utilizado en la sección anterior (GSE42952 [117]), el cual consta de tumores primarios y metástasis derivadas de cáncer de páncreas. La Figura 34 muestra el espacio de expresión génica resultante. Solo representamos la primera dimensión porque es la única significativa (el primer eje, eje X, explica el 39% de la varianza; $p = 0.13$).

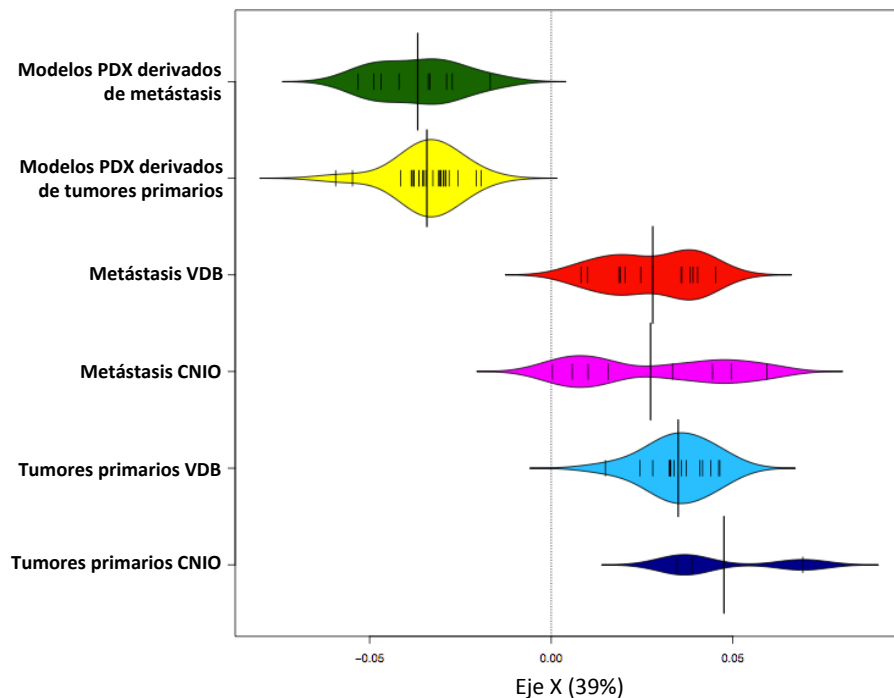
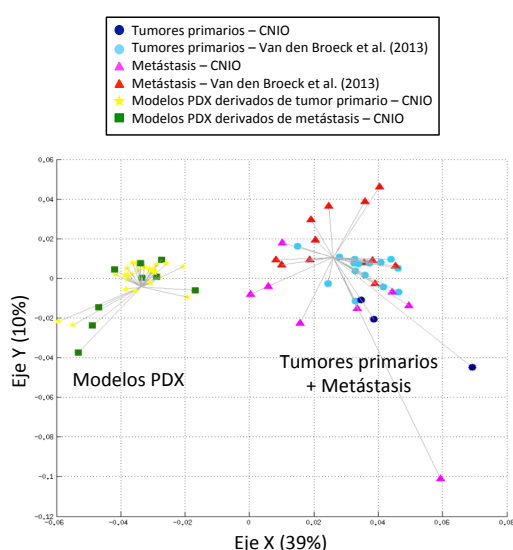


Figura 34. Coordenadas del eje X del espacio de expresión génica de tumores primarios y metástasis y los correspondientes modelos PDX (solo se representan las coordenadas X debido a que el segundo eje del MCA no es estadísticamente informativo, $p = 0.13$). El área coloreada representa la densidad de muestras en el espacio unidimensional, las líneas negras verticales y pequeñas corresponden a cada muestra, la línea negra más ancha de cada grupo representa el valor medio del mismo y la línea vertical punteada corresponde a la media global.

Una de las primeras observaciones en la Figura 34, es que los modelos PDX independientemente de su origen son iguales entre ellos y diferentes al resto de muestras (tumores primarios y metástasis obtenidos directamente del paciente). En la Figura 35 se muestra el mismo espacio de expresión que en la Figura 34 añadiendo el segundo eje, para facilitar la visualización de la distribución de los perfiles de expresión que pertenecen a un mismo individuo.

A) Tumores primarios, metástasis y modelos PDX



B) Tumores primarios, metástasis y modelos PDX según paciente

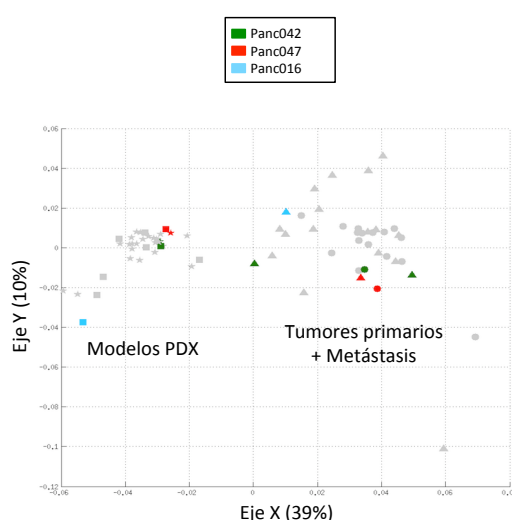


Figura 35. A) Espacio de expresión génica de tumores primarios y metástasis y los correspondientes modelos PDX (mismas muestras que la Figura 34). Cada punto representa el patrón de expresión de una muestra. El primer eje de separación es el eje X, el segundo es el eje Y, pero solo el primer eje es estadísticamente informativo ($p = 0.13$). Las líneas grises representan los grupos resultantes del análisis de *clustering* para $k = 2$ (ICH = 255.40). Cada línea une una muestra con el centro de masas del grupo al que pertenece. Cada tipo de muestra está representado con una forma diferente para facilitar la visualización. **B)** Espacio de expresión génica de tumores primarios y metástasis y los correspondientes modelos PDX (Figura 35A), representado según el paciente.

En la Figura 35B se observa que las muestras del mismo paciente están muy próximas en el espacio dentro de cada grupo principal (modelos PDX y muestras frescas humanas). Sin embargo, no se aprecia una aproximación de las muestras del mismo paciente entre los modelos PDX y los tumores primarios o metástasis. Estos resultados muestran que las muestras humanas frescas (tumor primario y metástasis) se agrupan juntas, al igual que los modelos PDX, independientemente de su origen.

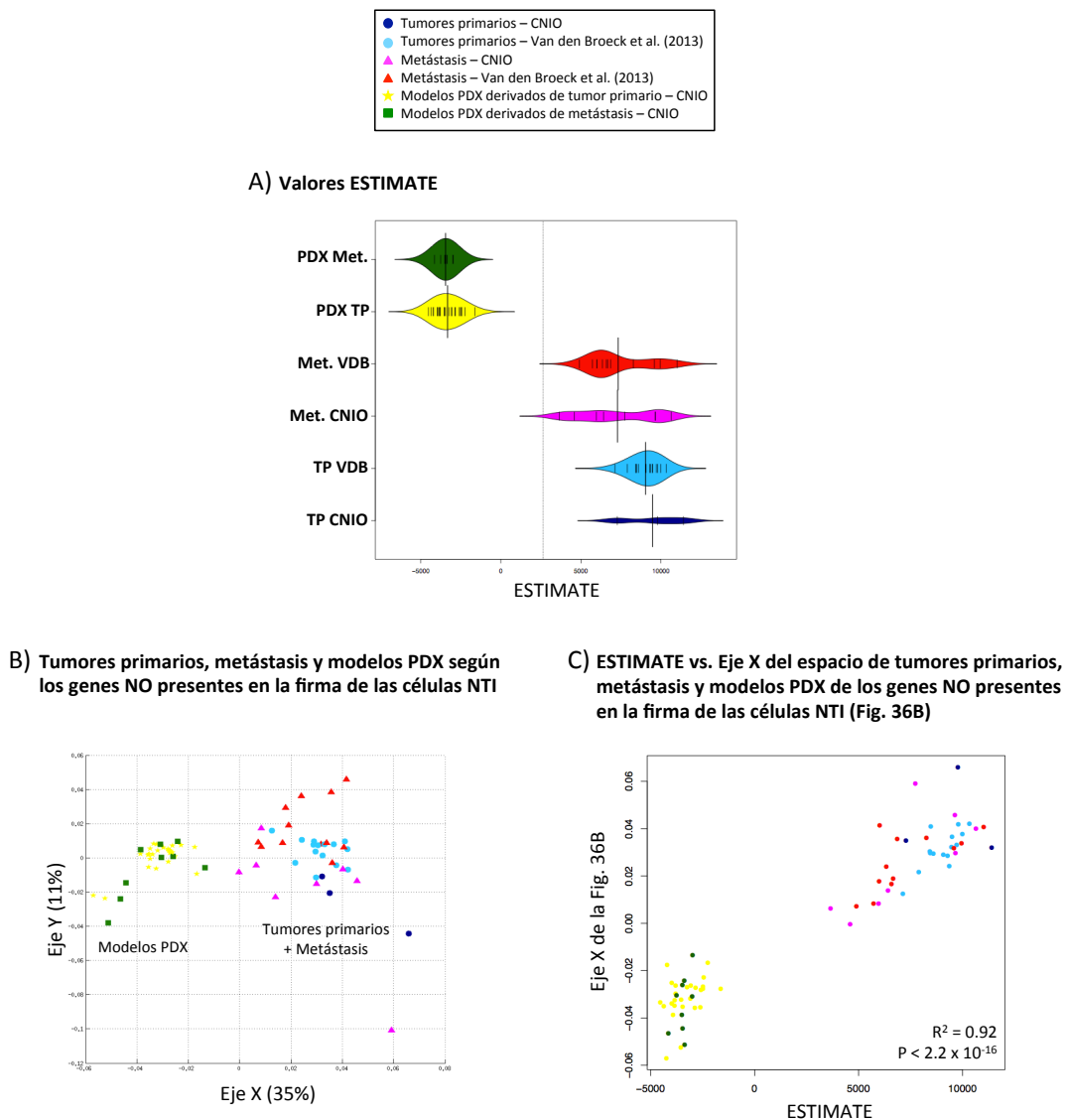


Figura 36. A) Valores ESTIMATE de los tumores primarios y metástasis y los correspondientes modelos PDX. Los valores altos indican una gran proporción de células NTI y viceversa. El área coloreada representa la densidad de muestras, las líneas negras verticales y pequeñas corresponden a cada muestra, la línea negra más

ancha de cada grupo representa el valor medio del mismo y la línea vertical punteada corresponde a la media global. **B)** Espacios de expresión génica de tumores primarios y metástasis y los correspondientes modelos PDX excluyendo las sondas de los genes presentes en la firma de las células NTI (mismas muestras que la Figura 34). Cada punto representa el patrón de expresión de una muestra. El primer eje de separación es el eje X, el segundo es el eje Y, pero solo el primer eje es significativo ($p = 0.13$). Solución óptima para $k = 2$ (ICH = 242.64). Cada tipo de muestra está representado con una forma diferente para facilitar la visualización. **C)** Correlación de los valores ESTIMATE respecto al eje X del espacio de expresión obtenido excluyendo las sondas de los genes presentes en la firma de las células NTI (Figura 36B).

Dados los resultados presentados en las secciones previas, en los cuales la separación de modelos PDX y de muestras humanas estaba causada por la presencia de células NTI, comprobamos si en este espacio de expresión génica el efecto de separación también se debe al estroma. Para ello, primero calculamos los valores ESTIMATE de estas muestras (Figura 36A). Los valores ESTIMATE de este resultado presentan una distribución muy similar a la de las muestras en el espacio unidimensional de la Figura 34. Lo cual apunta a que las células NTI, al igual que en los espacios anteriores, constituyen una fuente de variabilidad en la separación de las muestras en el espacio. A continuación, generamos el espacio de expresión sin las sondas pertenecientes a los genes de la firma de células NTI (Figura 36B), en el cual no se aprecian cambios significativos (al comparar este espacio con el de la Figura 35A). Finalmente, comprobamos la correlación del eje X del espacio de la Figura 36B con los valores ESTIMATE (Figura 36C). El valor de R^2 (0.92) y el valor p ($< 2.2 \times 10^{-16}$), sugieren que el efecto estroma también constituye una fuente fundamental de variabilidad en este caso.

Nuestro siguiente paso es la corrección *in silico* de dicho efecto, como hemos hecho en los espacios anteriores. La Figura 37 muestra la primera coordenada del espacio resultante (solo se muestra la primera dimensión porque es la única significativa en el espacio original, Figura 34). Tras la eliminación del efecto de las células NTI, se puede observar que los perfiles de expresión de las muestras son muy parecidos.

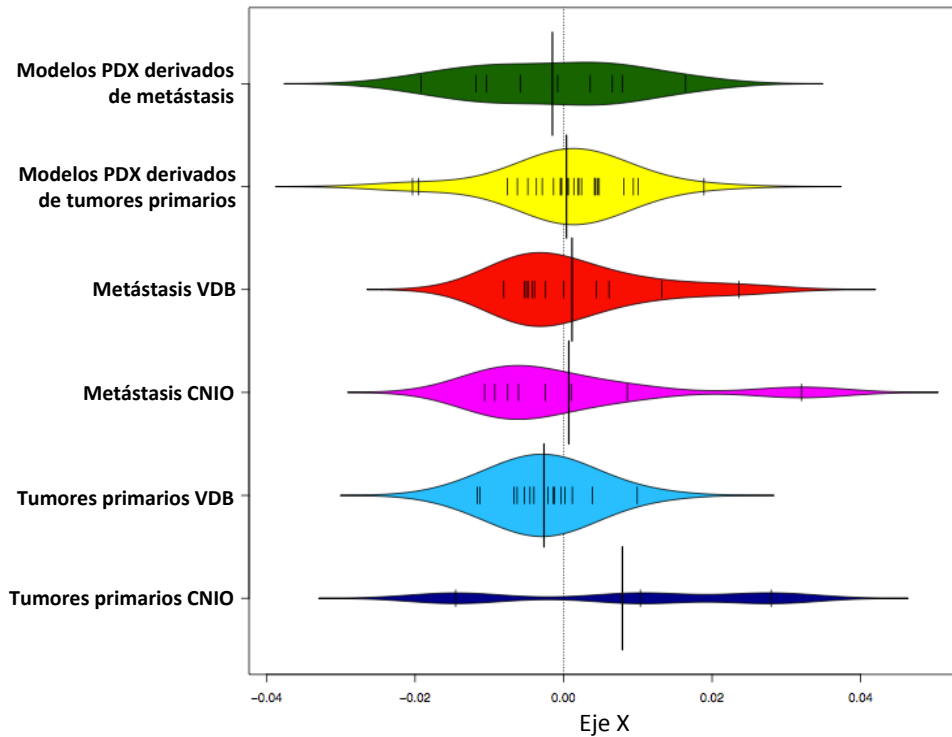


Figura 37. Coordenadas del eje X del espacio de expresión génica de tumores primarios y metástasis y los correspondientes modelos PDX tras la corrección (*in silico*) del efecto de las células NTI. Las coordenadas representadas corresponden a los residuos obtenidos de la regresión lineal del espacio de expresión génica de la Figura 36B (espacio obtenido excluyendo las sondas de los genes presentes en la firma de las células NTI). Solo se representan las coordenadas X debido a que el segundo eje del MCA de la Figura 36B, no es estadísticamente informativo ($p = 0.13$). El área coloreada representa la densidad de muestras en el espacio unidimensional, las líneas negras verticales y pequeñas corresponden a cada muestra, la línea negra más ancha de cada grupo representa el valor medio del mismo y la línea vertical punteada corresponde a la media global.

En resumen, todos estos resultados muestran que los perfiles de expresión de los modelos PDX derivados de metástasis son indistinguibles de los perfiles de los

modelos PDX derivados de tumor primario. Además, sugieren que la principal fuente de variación entre modelos PDX y las correspondientes muestras humanas es la presencia de células del estroma y sistema inmune. Cuando eliminamos esa fuente de variación, los perfiles de expresión de ambos tipos de modelos PDX, los tumores primarios y las metástasis solapan en el espacio. Esto avala que los modelos PDX son buenos representantes, a nivel transcripcional, de las muestras de las que proceden.

4.2.3. Respuesta a fármacos de los modelos PDX derivados de tumor primario y de metástasis

A partir del resultado de la sección anterior, que muestra que los modelos PDX son muy similares a nivel transcripcional independientemente de su origen, nos planteamos si ambos modelos poseen una capacidad predictora de sensibilidad a fármacos similar. Merece la pena evaluar esta posibilidad por su alto interés clínico, principalmente teniendo en cuenta que el 50% de los pacientes al momento del diagnóstico ya han desarrollado metástasis. Para abordar esta cuestión, diseñamos un experimento incluyendo los modelos PDX, tanto derivados de tumor primario como de metástasis, de los pacientes Panc042 y Panc047 (únicos pacientes de los que disponemos de ambos tipos de modelos PDX).

El experimento consta de tres ramas de tratamiento (Gemcitabina, Mitomicina C y Erlotinib) y una rama control (suero salino) (ver más detalles en la sección Materiales y métodos). La eficacia de los fármacos se ha analizado con 36 tumores en el modelo PDX derivado de tumor primario del paciente Panc042 (9 en cada rama de tratamiento) y 29 en el modelo PDX derivado de metástasis (7-8 en cada rama). En el caso del paciente Panc047 se han utilizado 24 tumores en el modelo PDX derivado de tumor primario (6 en cada rama de tratamiento) y 27 en el modelo PDX derivado de metástasis (6-7 en cada rama). En la Figura 38 se pueden observar los resultados obtenidos tras 4 semanas de tratamiento.

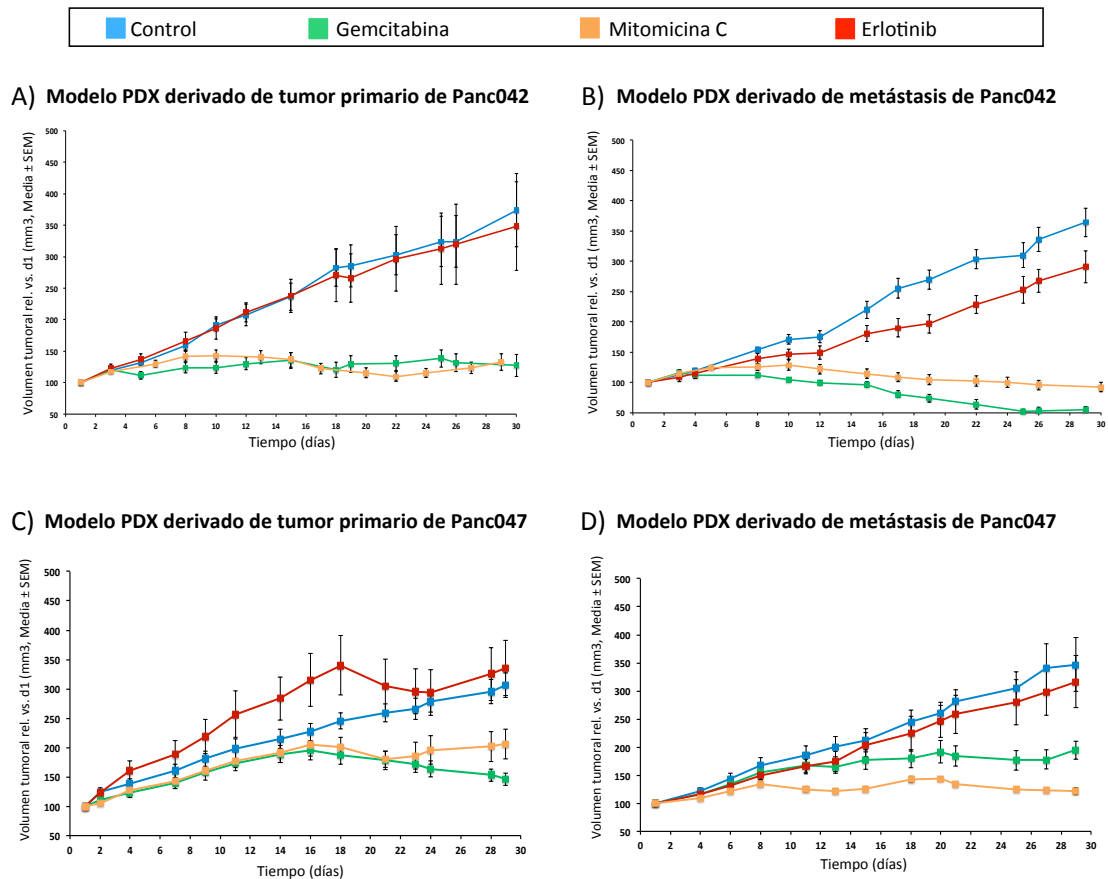


Figura 38. Sensibilidad a fármacos. **A)** Modelo PDX derivado de tumor primario del paciente Panc042. **B)** Modelo PDX derivado de metástasis del paciente Panc042. **C)** Modelo PDX derivado de tumor primario del paciente Panc047. **D)** Modelo PDX derivado de metástasis del paciente Panc047. Cada punto de la gráfica representa la media del volumen tumoral relativo (versus día 1) de los tumores pertenecientes a cada rama (6-9 tumores) hasta los 29-30 días de tratamiento. Las barras representan el error estándar medio (SEM, *Standard Error Mean*).

En la Figura 38 se puede observar que los modelos derivados de tumor primario y de metástasis, muestran una tendencia parecida en cuanto a su capacidad de respuesta (frente al conjunto de fármacos testado) en estos dos pacientes. Para poder cuantificar la similitud de ambos modelos en cada paciente, usamos el factor de inhibición del crecimiento tumoral (TGI, *Tumor Growth Inhibition*). Consideramos que los modelos PDX son respondedores o sensibles al tratamiento cuando el TGI sea superior al 90% (Tabla 6), ya que eso implica una reducción casi completa del tumor.

	%TGI = 1 - [(Tf - Ti) / (Cf - Ci)]		
	Gemcitabina	Erlotinib	Mitomicina C
Panc042 – Tumor primario	90%	6%	92%
Panc042 – Metástasis	118%	24%	102%
Panc047 – Tumor primario	75%	-16%	43%
Panc047 – Metástasis	56%	-2%	87%

Tabla 6. Factor de inhibición de crecimiento tumoral (TGI) de los modelos PDX derivados de tumor primario y metástasis. Código de colores: verde, sensibles, es decir, TGI > 90%; morado, TGI > 50%.

Los modelos PDX del paciente Panc042 son sensibles frente a Gemcitabina y Mitomicina C y resistentes a Erlotinib en ambos tipos de modelos, mostrando el patrón de similitud de respuesta esperado. Sin embargo, los modelos de Panc047 son resistentes para todos los fármacos testados. Curiosamente, si nos fijamos en los valores de la Tabla 6 (color morado), los modelos derivados de tumor primario y metástasis del paciente Panc047 también poseen valores más altos de TGI frente a Gemcitabina y Mitomicina C (TGI superiores a 50%). Por tanto, la tendencia observada parece seguir manteniéndose.

4.2.4. Predicción de los reguladores transcripcionales asociados a modelos PDX

Como hemos mencionado anteriormente, algunos perfiles de expresión pertenecen a un mismo paciente, correspondiendo a distintas localizaciones y tiempos en la evolución del tumor. Por ello, proponemos considerar los perfiles de expresión de dichas muestras, como parte de un proceso evolutivo y estudiar sus cambios a lo largo de distintos estadios (tumor primario, metástasis y modelo PDX).

La transcripción de un gen está regulada por la interacción de la RNA polimerasa con su promotor. En la mayoría de los genes, éste es el punto de control más importante, y

por ello, es el nivel más común de regulación. En la categoría de regulación se incluyen los promotores, las secuencias potenciadoras (*enhancers*) y las proteínas activadoras o inhibidoras (factores de transcripción, FT), que actúan mediante su unión a secuencias específicas de ADN (sitios de unión de factores de transcripción, TFBS, de sus siglas en inglés *Transcription Factor Binding Sites*). Debido a la importancia de la regulación de la transcripción en los procesos celulares, nos centramos en los cambios de actividad que se producen en los reguladores transcripcionales asociados a los TFBS. Para analizar dicha regulación, hemos usado el programa ISMARA (*Integrated System for Motif Activity Response Analysis*) [4], que nos permite estimar la actividad de FT conocidos en las muestras y compararlas. ISMARA se basa en las secuencias de TFBS, presentes en los promotores de los genes, y busca patrones de co-expresión en grupos de genes que comparten los mismos TFBS. Usando un modelo estadístico, el programa calcula una señal cuantitativa asociada a cada promotor del genoma y finalmente, en función de esa señal, predice que interacciones pueden estar regulando cada TFBS.

Dado que el principal modelo de estudio de esta tesis es el modelo PDX, nuestro primer paso es estudiar los perfiles de expresión de tumor primario y modelos PDX, para analizar que reguladores potenciales presentan cambios entre ambos. Utilizamos los datos de expresión de los pacientes Panc042 y Panc047 y planteamos la siguiente comparación: tumor primario vs. modelo PDX. ISMARA identifica 23 TFBS cuyos FT tendrían actividad diferencial ($z\text{-score} > 3$) en el paciente Panc042 y 22 en Panc047, siendo 5 comunes entre ambos individuos (TFBS regulados por los FT de las familias SNAI, NFATC, MYF, NFY y el factor EHF). Curiosamente, cuando estudiamos en estos mismos pacientes la ruta evolutiva: tumor primario → metástasis fresca humana → modelo PDX derivado de metástasis, los 5 TFBS alterados en los modelos PDX, también lo están en las muestras metastáticas (pertenecientes al humano o implantadas en los modelos PDX). En la Figura 39 podemos observar los 3 TFBS activados y los 2 desactivados, que presentan un actividad común entre los modelos PDX (comunes en los pacientes Panc042 y Panc047) y las metástasis o modelos PDX derivados de metástasis (comunes en ambos pacientes).

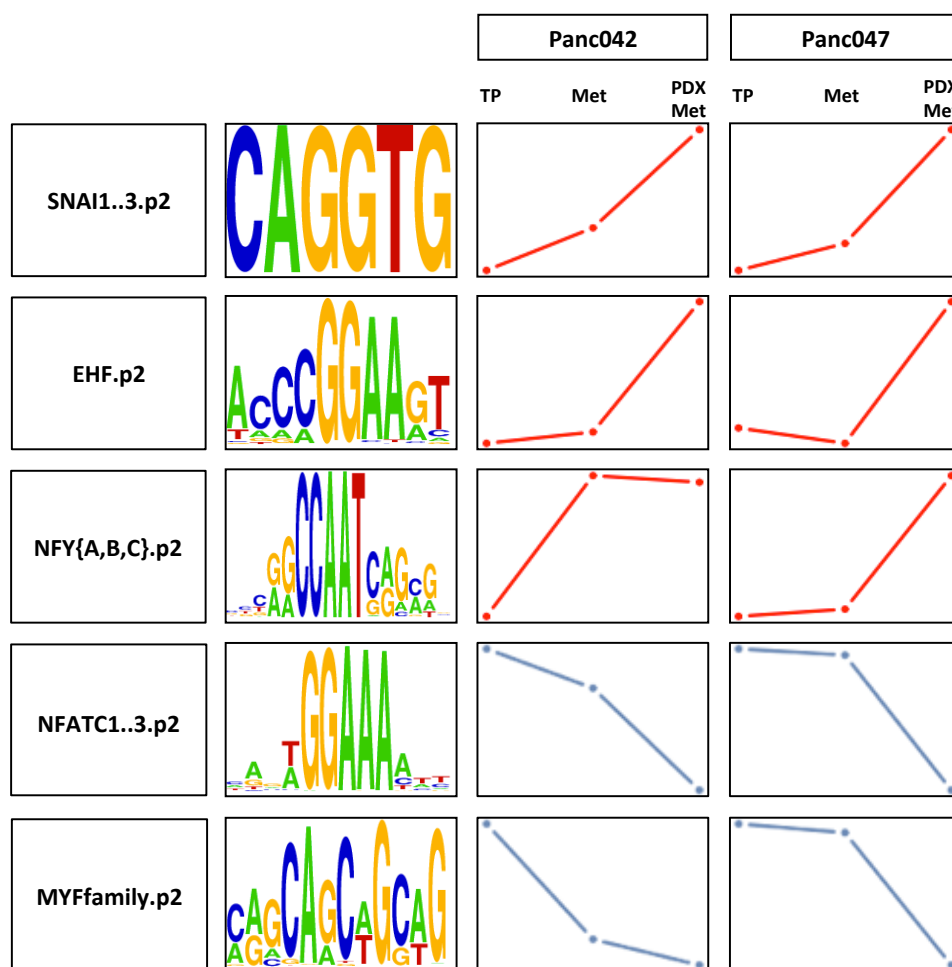


Figura 39. Motivos que presentan actividad diferencial significativa en la ruta evolutiva de tumor primario a modelo PDX y además en el proceso que implica a tumor primario (TP), metástasis fresca humana (Met) y modelo PDX derivado de metástasis (PDX Met). En la primera columna se indica el nombre del TFBS, en la segunda el logo con la secuencia a la que se unen los FT, y en la tercera y cuarta se muestran los perfiles de actividad en los pacientes Panc042 y Panc047, respectivamente. Código de color: rojo, aumento de actividad; azul, descenso de actividad.

SNAI1..3.p2 (TFBS regulado por la familia de FT SNAI (*snail family zinc finger*), formada por SNAI1, SNAI2 y SNAI3), EHF.p2 (regulado por el factor EHF, *Ets Homologous Factor*) y NFY{A,B,C}.p2 (regulado por los FT de la familia NFY (*Nuclear transcription Factor Y*): NFYA, NFYB, NFYC) son los TFBS que muestran

un aumento de actividad en ambas rutas evolutivas. Mientras que NFATC1..3.p2 (regulado por los FT de la familia NFATC (*Nuclear Factor of Activated T-cells, Cytoplasmic*): NFATC1, NFATC2 y NFATC3) y MYFfamily.p2 (regulado por los FT de la familia MYF (*Myogenic Factor*): MYOD1, MYF6, MYF5 y MYOG) se encuentran reprimidos. Estos FT tienen funciones diversas: los factores SNAI y EHF se encuentran implicados en la transición epitelio-mesénquima [1, 63], los NFY están relacionados con ciclo celular [96], los NFATC están implicados en la activación de las células T del sistema inmune [72] y los FT de la familia MYF son factores miogénicos relacionados con la diferenciación muscular [123].

Aunque el reducido número de muestras sigue siendo un problema, estos resultados nos permiten observar que existen diferentes TFBS que presentan alteraciones de actividad en los modelos PDX derivados de tumores primarios y de metástasis. En alguno de ellos, además parece que la muestra de metástasis representa un estado intermedio entre el tumor primario y el modelo PDX. Por ello, es recomendable mantener el esfuerzo de conseguir muestras del mismo paciente que nos permitan seguir analizando la actividad diferencial de los FT, dado que pueden constituir posibles dianas terapéuticas.

5. Discusión

5.1. Modelos PDX de ADP derivados de tumor primario

5.1.1. El modelo PDX como modelo de investigación y para estudios preclínicos

Los modelos PDX tienen numerosas ventajas frente a los modelos clásicos utilizados en estudios preclínicos (líneas celulares y xenoinjertos derivados de líneas celulares), pero también tienen limitaciones que no deben ser pasadas por alto. El desarrollo de un modelo de xenoinjerto derivado de paciente requiere un gran esfuerzo humano y económico. Es necesaria una excelente coordinación entre muchos equipos de diversas áreas, desde cirugía y anatomía patológica en el hospital, hasta el laboratorio donde se lleva a cabo el injerto. Además, es necesaria la infraestructura adecuada para el mantenimiento de los animales (animalario) y un equipo cualificado de técnicos y veterinarios que se encarguen de vigilar que los animales están en las condiciones adecuadas, para que se cumplan todas las consideraciones éticas y legales necesarias para el bienestar de los animales durante los experimentos. Todo ello implica un coste de mantenimiento muy elevado, por ejemplo, la generación de un modelo PDX tiene un coste aproximado de 8000 euros, incluyendo la reimplantación hasta el pase número 4, en el cual ya se considera que el modelo PDX ha sido perpetuado. Además, en un estudio de eficacia de fármacos, cada rama de tratamiento supone aproximadamente unos 1000 euros.

Los modelos PDX que se han utilizado en el presente estudio, han sido siempre implantados de manera subcutánea. Hemos demostrado en este trabajo que, en lo que se refiere a perfiles de expresión, los modelos implantados subcutánea y ortotópicamente no presentan diferencias que permitan distinguirlos. Un trabajo realizado con ambos tipos de modelos procedentes de un panel de diversos tipos de cáncer (cáncer de mama, gastrointestinal, genitourinario, ginecológico, de pulmón, muscular, neurológico, de cabeza y cuello, en la piel, linfoma y leucemia) llegó a conclusiones similares [84]. Asimismo, nuestros resultados coinciden con los de

Mattie et al. (2013) [79], los cuales mostraron ya una alta correlación entre modelos PDX implantados subcutánea y ortotópicamente procedentes de ADP.

Sin embargo, no se puede olvidar que en el modelo PDX implantado subcutáneamente se pierde la interacción con el ambiente pancreático. Se sabe que los modelos PDX ortotópicos tienen un crecimiento tumoral más robusto que los subcutáneos [9]. Sin embargo, no tenemos suficiente información para asegurar que esa pérdida no esté afectando a otros aspectos importantes del modelo, como por ejemplo la distribución o difusión de un fármaco o la interacción entre proteínas humanas producidas por el tumor y proteínas murinas. Por ello, es necesario intensificar los esfuerzos en caracterizar el modelo PDX implantado subcutáneamente y su idoneidad como modelo preclínico y de investigación, en los ámbitos en los cuales no se ha llevado a cabo todavía. Algunos autores incluso proponen que la situación ideal de un estudio preclínico, estaría compuesta por una combinación de los diferentes tipos de modelos animales (subcutáneos, ortotópicos y genéticamente modificados) en función de la pregunta biológica específica de cada experimento [79].

5.1.2. Los modelos PDX mantienen una memoria de expresión génica de los tumores de origen

Uno de los primeros trabajos en analizar un espacio de expresión génica fue el de Lukk et al. (2010) [70]. En él generaron un mapa global de expresión génica utilizando líneas celulares, muestras de tejidos sanos y muestras tumorales de diferentes órganos. En ese estudio se demostró que las líneas celulares poseen una alta similitud entre ellas, a pesar de sus diferentes orígenes [70]. Igualmente, este trabajo evidenció que las células tumorales son un estado intermedio entre las células de tejidos sanos y las líneas celulares inmortalizadas. De forma adicional, se han publicado diversos trabajos, basados en perfiles de expresión, que demuestran que los modelos preclínicos de xenoinjertos basados en líneas celulares sufren cambios en los perfiles de expresión, lo que a su vez provoca un alejamiento de los patrones de expresión del tumor original [20, 28, 84].

Nuestras observaciones, basadas en los espacios de expresión génica utilizando tumores primarios en combinación con modelos PDX o líneas celulares, son relevantes porque demuestran que los modelos PDX mantienen, en cierta medida, una memoria de expresión de su tejido de origen, mientras que las líneas celulares son tan parecidas entre ellas que es imposible determinar su origen. Este escenario nos permite remarcar la influencia del ambiente celular en la expresión global de las muestras de tejido primario humano, líneas celulares y modelos PDX. Nuestros resultados son consistentes con varios estudios previos realizados usando modelos PDX de ADP [79], cáncer de pulmón microcítico o de células pequeñas [20], melanoma uveal [64], mama [22] y riñón [103]. En todos estos estudios, se ha observado un mantenimiento de las características claves del tumor original, incluyendo la actividad funcional y los perfiles de expresión génica. De manera adicional, Daniel et al. (2009) [20] evidenciaron que las divergencias genéticas entre tumores originales y líneas celulares eran mayores que las divergencias genéticas entre tumores primarios y modelos PDX. Ese hecho es confirmado por nuestros resultados.

5.1.3. Las células no tumorales infiltrantes contribuyen a la subestimación de las similitudes entre los modelos PDX y los tumores primarios

Las células no tumorales infiltrantes (NTI) pertenecen principalmente al estroma y al sistema inmune. En los tumores injertados en modelos PDX, el estroma humano es sustituido por estroma murino y su cantidad relativa disminuye respecto al tumor original. Nuestros resultados muestran que hay una fuerte correlación entre las diferencias transcripcionales de modelos PDX y tumores primarios, en función de los diferentes porcentajes de estroma. Esto implica que las diferencias transcripcionales observadas entre modelos PDX y tumores primarios, están siendo sobreestimadas en los análisis de expresión génica. Debido a ello, las diferencias entre ambos grupos son atribuibles, en su mayor medida, a la diferente contribución de las células NTI. De hecho, cuando el efecto de las células NTI es corregido en el espacio de expresión génica, los modelos PDX y los tumores primarios son prácticamente indistinguibles.

Nuestros datos sugieren que las células NTI murinas, presentes en los modelos PDX, tienen una contribución menor en los perfiles de expresión cuando se usa la plataforma *GeneChip® Human Genome U133 Plus 2.0 Array*. En este sentido, otros estudios previos han tratado de determinar la especificidad de especie de las plataformas de hibridación humanas en muestras con mezcla de células humanas y murinas. La conclusión general fue que la mayor parte de la señal detectada en estas plataformas es específica de humano [10, 88, 100]. Asimismo, las nuevas técnicas de secuenciación de ARN (*RNAseq*) permiten la cuantificación simultánea de secuencias de lectura específicas de diferentes tipos celulares. Cuando esos tipos celulares son células de ratón (procedentes del estroma) y de humano (pertenecientes al tumor), estas técnicas constituyen una gran herramienta para dilucidar la contribución del estroma murino en los cambios de expresión [10, 17, 116].

Una de las propuestas para eliminar el efecto de las células NTI contaminantes en los diferentes tipos celulares, es la microdissección. Esta tecnología debería constituir la mejor forma de determinar la contribución real de cada tipo celular en el perfil de expresión de un tejido. Nuestros resultados sugieren que las muestras microdisseccionadas no representan adecuadamente al tumor primario completo y que, de hecho, los modelos PDX son más parecidos a los tumores primarios originales. En los espacios de expresión observamos que parte de esas diferencias se deben, en cierta medida, a la detección del estroma en las muestras no microdisseccionadas, pero también sugieren que existe una fuente de variabilidad probablemente inherente a la técnica. Esto se puede deber a que la microdissección celular implica la fijación del tejido, la cual ha sido previamente asociada con altos niveles de degradación del ARN. Asimismo, durante el procesamiento de las muestras se puede producir la amplificación de fragmentos de ARN parcialmente degradados. Ambos efectos, podrían estar provocando sesgos en los niveles de expresión [36, 88].

Es importante tener en cuenta que los perfiles de expresión de los tumores primarios microdisseccionados, fueron combinados con datos de tumores primarios completos, no microdisseccionados [3], y que dicha combinación fue usada para la definición de subtipos [16]. En nuestros análisis no hemos podido identificar los subtipos, ni al

utilizar todos los genes ni cuando usábamos solo los genes de la firma PDAsgner definida por los autores. Por ello, creemos que es fundamental tener cuidado al interpretar los resultados de muestras microdisseccionadas. Los resultados disponibles sugieren que no es recomendable usar conjuntamente perfiles de expresión de muestras microdisseccionadas con muestras de tejido completo, ya que son muy diferentes.

5.1.4. Las alteraciones funcionales se producen por la adaptación al microambiente murino y se mantienen estables a lo largo de los pases

A pesar de las ventajas de los modelos PDX sobre las líneas celulares, los tumores humanos injertados en ratón muestran diferencias respecto al tumor primario original [113]. Por ello, se debe tener precaución a la hora de interpretar los resultados obtenidos al usar estos modelos. Los resultados de esta tesis muestran que los perfiles de expresión de los modelos PDX derivados de ADP y de CHC conservan parte de los patrones de expresión de sus correspondientes tejidos de origen. Sin embargo, también hemos identificado varias rutas biológicas cuyos genes muestran una expresión diferencial entre modelos PDX y tumores primarios.

Las rutas biológicas que se encuentran alteradas en los modelos PDX están relacionadas con ciclo celular, replicación del ADN, transducción de señales, hemostasia y organización de la matriz extracelular. Los genes implicados en los dos primeros grupos funcionales, ciclo celular y replicación del ADN, están principalmente relacionados con los subgrupos de segregación cromosómica y regulación de la división celular. La sobre-expresión de estos genes coincide con las altas tasas de proliferación observadas en los modelos PDX al ser injertados en el ratón [30]. Los grupos funcionales restantes, transducción de señales, hemostasia y organización de la matriz extracelular, contienen genes reprimidos en modelos PDX, incluso algunos son comunes en más de un grupo. La disminución de los niveles de expresión de estos genes en los modelos PDX, puede ser una consecuencia de la eliminación de las células NTI de tumor primario humano tras el injerto, mediante el

reemplazo de las células del estroma humano por elementos murinos, como ya han sugerido recientemente otros autores [79]. Nuestros resultados sugieren que las células NTI contribuyen, en gran manera, a las diferencias entre modelos PDX y tumores primarios, pero también hay que tener en cuenta otros posibles efectos. Por ejemplo, el hecho de que las células tumorales humanas podrían responder de manera poco eficiente a los estímulos del estroma del ratón. En cierta medida, la expresión diferencial de estos genes está reflejando la adaptación del tumor al microambiente murino. De hecho, la mayoría de las funciones alteradas han sido previamente descritas en la misma dirección en modelos PDX de melanoma uveal [64]. Algunas de ellas también se han encontrado alteradas en modelos PDX derivados de cáncer de páncreas [79] y de cáncer de mama [95].

Adicionalmente, también hemos podido demostrar que las alteraciones en los grupos funcionales de los modelos PDX, se producen en los estadios iniciales de la adaptación al nuevo ambiente, tras el implante, y que posteriormente no se producen cambios significativos. En otras palabras, los patrones de expresión son estables a lo largo de los sucesivos pases. Este resultado concuerda con estudios realizados previamente en páncreas [79] y en otros tipos tumorales, como el cáncer de pulmón microcítico [20], el melanoma uveal [64], el cáncer de mama [95] o el cáncer colorectal [114], incluso en un estudio con un panel de diversos tipos tumorales [84]. Este resultado es muy relevante porque confirma que además de la estabilidad del estatus mutacional, demostrada previamente [79, 98, 120], podemos demostrar que existe una estabilidad transcripcional en los modelos PDX derivados de ADP. Por tanto, estos resultados avalan la utilidad del modelo PDX como modelo de investigación y preclínico, sobre todo en enfermedades tan agresivas como el ADP.

5.2. Modelos PDX de ADP derivados de metástasis

5.2.1. Los modelos PDX y las metástasis comparten alteraciones funcionales

Las rutas biológicas alteradas en los modelos PDX están, además, típicamente desreguladas en las muestras metastáticas comparadas con los tumores primarios. De hecho, algunos autores han propuesto que los modelos PDX imitan algunas características metastáticas [30, 60, 88, 109]. El tumor injertado podría, por tanto, encontrarse en una situación de “metástasis forzada”, ya que el tumor necesita colonizar y sobrevivir en un nuevo ambiente. Este fenotipo metastático presente en los modelos PDX, es coherente con nuestros resultados, donde la proliferación se encuentra sobre-activada y los procesos de señalización e interacción con la matriz extracelular se han visto reprimidos. Para comprobar si existe una similitud, a nivel funcional, entre modelos PDX y metástasis respecto a los tumores primarios, hemos analizado dos conjuntos de datos públicos. Aunque había algunas diferencias dependiendo del tejido de colonización, las rutas biológicas alteradas en los modelos PDX estaban significativamente alteradas en las metástasis. Un tumor primario está constituido por una población de diferentes clones, siendo los clones metastáticos solo una parte de la población total del tumor [118]. Las células que injertan de manera exitosa en un ratón, probablemente proceden de una de esas subpoblaciones clonales con potencial metastático. Obviamente, son necesarios estudios adicionales para mejorar el entendimiento de los mecanismos implicados en los procesos de colonización de las células tumorales en los nuevos nichos, ya sea de forma natural en un paciente con metástasis o de manera artificial en un modelo PDX. Sería particularmente útil entender si el efecto de tropismo, o lo que es lo mismo que la metástasis se produzca en una localización determinada, puede contribuir a las diferencias en los perfiles de expresión génica.

5.2.2. Los modelos PDX derivados de tumor primario son indistinguibles de los modelos PDX derivados de metástasis

A nivel funcional, los modelos PDX y las metástasis frescas comparten alteraciones, como hemos demostrado en la sección anterior. Previamente hemos mencionado la existencia de varios trabajos que muestran la estabilidad genómica de los modelos PDX respecto a sus tumores de origen, en los cuales se estudia el número de copia y los niveles de expresión de los genes [64, 75]. Los modelos PDX de esos trabajos derivan indistintamente de tumores primarios o metástasis, sugiriendo que ambos modelos son parecidos a dichos niveles. Por ello, a continuación hemos comprobado las similitudes transcripcionales existentes entre los modelos PDX derivados de tumor primario o de metástasis. Nosotros hemos generado modelos PDX derivados de metástasis, para comprobar la distribución en un espacio de expresión génica respecto al resto de muestras. Nuestros resultados ponen de manifiesto que los modelos PDX derivados de metástasis son indistinguibles de los modelos PDX derivados de tumor primario, apoyando los resultados previos.

5.2.3. Los modelos PDX derivados de tumor primario y los modelos PDX derivados de metástasis parecen mostrar un sensibilidad a fármacos similar

En un estudio reciente, desarrollado por Kerbel et al. (2013), se usaron fármacos de primera línea habituales tras la cirugía o fármacos más apropiados para estadios avanzados (metástasis), en modelos PDX derivados de metástasis de pacientes con cáncer de mama [60]. En este estudio, resaltan el uso de modelos PDX como modelos preclínicos como una estrategia muy prometedora, debido a que la mayoría de los modelos PDX utilizados en sus experimentos, mimetizan la respuesta frente a determinados fármacos usados habitualmente para tratar metástasis, pero también para aquellos usados para tratar tumores primarios tras la cirugía.

Los modelos PDX parecen representar una situación de metástasis forzada y, además hay indicios de que son capaces de mimetizar y predecir la sensibilidad a un fármaco frente a enfermedad resecable o metastásica. Por ello, a continuación, diseñamos un experimento para comprobar si la sensibilidad frente a un conjunto de fármacos, en los modelos PDX de ADP derivados de tumor primario y metástasis, era similar. Cada paciente puede responder de manera diferente frente a un fármaco, por ello consideramos necesario que las muestras de tumor primario y metástasis, a partir de las cuales se iban a generar los modelos PDX, derivaran del mismo paciente. Para ello, aprovechamos el programa establecido por el Dr. Manuel Hidalgo en los hospitales del Grupo Hospital de Madrid, para llevar a cabo “autopsias calientes”, en las cuales se extrae muestra de todas las lesiones cancerígenas (tumores primarios y metástasis) que presente un paciente de oncología en el momento de su fallecimiento. A pesar del esfuerzo realizado por todas las partes implicadas (cirujanos, patólogos, veterinarios y técnicos de laboratorio y animalario) en la obtención de muestras, en estos últimos 3 años desde que comenzó este programa, solo hemos conseguido que ambos tipos de muestras injertaran con éxito a partir de dos pacientes. En cualquier caso, creemos que la tendencia observada en estos resultados preliminares es muy prometedora, y es informativa de la utilidad del experimento para continuarlo en el futuro.

Los modelos PDX de uno de los pacientes, Panc042, mostraron una respuesta prácticamente idéntica para todos los fármacos testados. Adicionalmente, los modelos PDX del otro paciente, Panc047, mostraron también una tendencia similar en las respuestas a todo el conjunto de fármacos, si bien no se alcanzó el nivel de reducción de tumor necesario para considerar la existencia de sensibilidad. Si tenemos en cuenta el TGI, medida que nos permite cuantificar la respuesta frente a cada fármaco, podemos ver que en ambos modelos PDX, independientemente de su origen, las respuestas a los tres fármacos testados poseen una tendencia similar. Hubiera sido necesario alargar el experimento del paciente Panc047, para analizar si la sensibilidad se hubiera alcanzado tras unos días extra de tratamiento.

Como hemos mencionado previamente, con solo dos pacientes no tenemos potencia estadística para obtener conclusiones generales a partir de este experimento, pero los resultados parecen indicar la existencia de una respuesta similar frente a diversos fármacos, independientemente de si el modelo PDX deriva de tumor primario o de metástasis. Esta tendencia es coherente con los resultados del estudio de Kerbel et al. (2013) [60], mencionado al principio de esta sección.

Nuestros resultados también podrían explicarse según las observaciones del trabajo de Yachida et al. (2010) [124], en el cual se detectó que los clones presentes en las metástasis de pacientes con cáncer de páncreas, ya estaban presentes en los correspondientes tumores primarios. En concreto, se demostró que las mutaciones somáticas presentes en las metástasis, estaban en su mayoría presentes en el clon parental del tumor primario. Yachida et al. propusieron la existencia de una evolución clonal, en la cual se partiría de un clon con un elevado número de alteraciones. Las células descendientes irían adquiriendo nuevas mutaciones, menos numerosas pero que determinarían la capacidad de colonizar en un órgano u otro. En la actualidad, estamos llevando a cabo una colaboración con la Unidad de Bioinformática Traslacional del CNIO, para analizar las mutaciones somáticas presentes en las muestras humanas (tumor primario y metástasis) y los modelos PDX (derivados de dichas muestras humanas) de ambos pacientes.

5.2.4. Factores de transcripción con actividad diferencial en modelos PDX

La existencia de cambios en los patrones de expresión de los modelos PDX respecto de los tumores primarios, nos hace pensar en los reguladores transcripcionales que serían responsables de esos cambios. Dado que disponíamos, para dos pacientes, de muestras de tumor primario, metástasis y modelos PDX derivados de ambos, hemos analizado los perfiles de expresión como parte de un “proceso evolutivo” (las muestras proceden de distintas localizaciones y tiempos en la evolución del tumor). Además, hemos buscado los posibles factores de transcripción (FT) reguladores de las distintas rutas evolutivas.

Una regulación adecuada de la transcripción es fundamental para el correcto desarrollo celular. La actividad o expresión aberrante de los FT da lugar a un crecimiento celular impropio, desembocando en enfermedades como el cáncer [11]. Esto se debe a que la ganancia o pérdida de función de los FT puede transformarlos en oncogenes o genes supresores de tumores. La actividad en los sitios de unión de factores de transcripción (TFBS, *Transcription Factor Binding Sites*) está, por tanto, implicada en la progresión del tumor, pero también afecta a la metástasis [25]. De hecho, muchos tipos tumorales dependen de la actividad aberrante de los FT, principalmente de aquellos que adquieren actividad oncogénica. La inhibición directa de la expresión de estos motivos, ha demostrado tener actividad antitumoral. También se están estudiando nuevas estrategias centradas en impedir su interacción con otros FT o incluso restringir su acceso a la cromatina. Por ello, los FT son importantes dianas terapéuticas contra el cáncer [125].

Inicialmente, estudiamos la regulación transcripcional de la ruta desde tumor primario a modelo PDX. El hecho de encontrar patrones alterados en los modelos PDX respecto a los tumores primarios, es coherente con los resultados hallados en los espacios de expresión génica, donde observábamos ciertas diferencias entre ambos tipos de muestras. Por último, nos centramos en si los cambios que suceden de tumor primario a modelo PDX, también están presentes en metástasis y en el modelo PDX derivado de metástasis.

Lo más interesante de este análisis es que todos los FT que presentaban una actividad diferencial en los modelos PDX derivados de tumor primario, continúan mostrándola en las metástasis y los modelos PDX derivados de las mismas. Los FT que presentan actividad diferencial son muy diversos. Algunos cambios en los patrones de actividad, pueden estar causados por la adaptación al microambiente murino, al igual que lo estaban las rutas biológicas alteradas en los modelos PDX respecto a los tumores primarios. Por ejemplo, la desactivación propuesta de los factores de la familia NFATC (implicados en la activación de las células T del sistema inmune [72]). Aunque el estroma murino secreta los elementos necesarios para estimular el sistema inmune del ratón, éste es atímico por lo que su sistema inmune es deficiente y

tampoco puede producir todas las células necesarias para una activación completa del sistema inmunitario. La desactivación de NFATC estaría, por tanto, condicionada a este hecho. También hemos detectado un aumento de actividad de los FT de la familia SNAI y el factor EHF (relacionados con la transición epitelio-mesénquima, lo cual favorece la migración de las células tumorales por el torrente sanguíneo hasta la localización donde darán lugar a metástasis [1, 63]). El aumento de actividad de estos FT puede estar condicionado por el nuevo ambiente, debido a la situación forzada de metástasis a la que es sometida el injerto. Nuestros resultados sugieren que el ambiente murino causa determinados cambios en las células tumorales, tanto a nivel de la expresión génica como en la actividad de los FT.

En resumen, tenemos varios FT alterados en los modelos PDX derivados de tumor primario y que continúan alterados en modelos PDX derivados de metástasis, lo cual indica que hay una adaptación al medio murino. Sin embargo, la falta de muestras del mismo paciente representa una limitación de este trabajo, dado que solo hemos tenido dos pacientes para crear rutas evolutivas a estudiar. Por tanto, nuestros resultados acerca de la actividad diferencial de los FT deben considerarse preliminares.

A lo largo de esta tesis, hemos mostrado que los modelos PDX son buenos modelos preclínicos y de investigación, ya que mantienen parte de la expresión de su tumor de origen. Parte de las diferencias entre ambos tipos de muestra, es debida a la presencia de células no tumorales infiltrantes detectadas en los perfiles de los tumores primarios y no en los modelos PDX. Los cambios transcripcionales que se producen, lo hacen por la adaptación al nuevo medio y se mantienen estables a lo largo de los pases. Dichos cambios en los niveles de expresión junto a los cambios detectados en sus reguladores, soportan nuestra hipótesis de que los modelos PDX se comportan como las metástasis, aumentando la actividad en los procesos de proliferación y disminuyendo los relacionados con la comunicación. Además, esta tesis da como resultado una metodología robusta y cuantitativa para el análisis de perfiles de expresión en espacios de expresión génica.

En el futuro, será necesario continuar la caracterización de los modelos PDX para profundizar en el conocimiento de la biología de los mismos, y aprender todo lo posible sobre su utilidad y limitaciones como modelos preclínicos de cáncer. Será fundamental ampliar nuestros resultados sobre los cambios de expresión, así como continuar los estudios sobre los elementos reguladores, a partir de la información preliminar obtenida en esta tesis, y combinarlos con nuevos análisis de epigenética y proteómica. Este esfuerzo es muy importante en el campo del ADP, pero también en otros tipos de cáncer, ya que eso nos permitiría llevar a cabo análisis comparativos que permitan identificar mecanismos específicos del ADP o comunes con otros tipos de cáncer.

El cáncer es una enfermedad genéticamente heterogénea a nivel inter- e intra-individuo. El modelo PDX representa una visión ligeramente estática de un proceso altamente dinámico, y es posible que no pueda reproducir toda la diversidad del ADP. Por ello, creemos que es fundamental para la investigación sobre el cáncer, en general, y del ADP, en particular, conseguir más muestras de tumor primario y metástasis de los pacientes. La incorporación de nuevas muestras permitirá generar nuevos modelos PDX, ampliando así el biobanco del CNIO y su diversidad. Además, un mayor número de muestras nos permitiría aumentar la potencia estadística y realizar análisis mucho más robustos y exhaustivos de dichos tumores. En este sentido, recientemente se ha creado un consorcio, denominado *EurOPDX Consortium* [40], que se centrará en la creación de una red global con toda la información asociada a los modelos PDX generados por los distintos grupos que lo componen (información clínica de los pacientes, perfiles moleculares de expresión, alteraciones en el número de copia, proteómica, farmacogenómica...). Esta iniciativa tiene una gran relevancia clínica ya que permitirá mejorar las limitaciones que presentan actualmente los modelos PDX. Adicionalmente, el *EurOPDX Consortium* pretende tener un impacto real en la calidad de vida de los pacientes con cáncer, así como en su supervivencia, gracias a ensayos preclínicos multicentro en Europa.

6. Conclusiones

1. El ambiente murino afecta al tejido humano injertado en ratón provocando alteraciones en el transcriptoma. Los modelos PDX todavía conservan importantes aspectos de los perfiles de expresión de sus tejidos de origen. Esto no sucede en las líneas celulares.
2. Las células no tumorales infiltrantes constituyen una contaminación detectada en los perfiles de expresión de los tumores primarios, la cual desemboca en una sobrestimación de las diferencias entre dichas muestras y los modelos PDX. Adicionalmente, los modelos PDX representan mejor al tumor primario que las muestras de tumor microdisseccionadas.
3. Los procesos de comunicación con el ambiente murino se encuentran reprimidos en los modelos PDX, mientras que los procesos asociados a proliferación están sobre-expresados. Este estado funcional implica que los modelos PDX representan al tumor primario en cierta medida, pero también muestra que sufren cambios adaptativos frente al nuevo ambiente.
4. Aunque algunas rutas biológicas sufren cambios tras la implantación del tumor en el ratón, esos cambios permanecen estables a lo largo de los pases. Desde nuestro punto de vista, la estabilidad de la expresión génica en los modelos PDX de ADP es muy importante, ya que refuerza su utilidad como modelos preclínicos y de investigación de la enfermedad.
5. Con los datos disponibles, los modelos PDX derivados de tumor primario y los modelos PDX derivados de metástasis, son indistinguibles a nivel transcripcional y muestran similitudes en sus respuestas frente a un conjunto determinado de fármacos. Estos resultados apuntan a que los modelos PDX pueden representar una situación de metástasis forzada.

BIBLIOGRAFÍA

1. Albino D, Longoni N, Curti L, Mello-Grand M, Pinton S, Civenni G, Thalmann G, D'Ambrosio G, Sarti M, Sessa F, Chiorino G, Catapano CV, Carbone GM. **ESE3/EHF controls epithelial cell differentiation and its loss leads to prostate tumors with mesenchymal and stem-like features.** *Cancer Res* 2012, **72**:2889-2900.
2. Allan AL, Vantyghem SA, Tuck AB, Chambers AF. **Tumor dormancy and cancer stem cells: implications for the biology and treatment of breast cancer metastasis.** *Breast Dis* 2006, **26**:87-98.
3. Badea L, Herlea V, Dima SO, Dumitrascu T, Popescu I. **Combined gene expression analysis of whole-tissue and microdissected pancreatic ductal adenocarcinoma identifies genes specifically overexpressed in tumor epithelia.** *Hepatogastroenterology* 2008, **55**:2016-2027.
4. Balwierz PJ, Pachkov M, Arnold P, Gruber AJ, Zavolan M, van Nimwegen E. **ISMARA: automated modeling of genomic signals as a democracy of regulatory motifs.** *Genome Res* 2014, **24**:869-884.
5. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, Wilson CJ, Lehár J, Kryukov GV, Sonkin D, Reddy A, Liu M, Murray L, Berger MF, Monahan JE, Morais P, Meltzer J, Korejwa A, Jané-Valbuena J, Mapa FA, Thibault J, Bric-Furlong E, Raman P, Shipway A, Engels IH, Cheng J, Yu GK, Yu J, Aspesi P, de Silva M, Jagtap K, Jones MD, Wang L, Hatton C, Palescandolo E, Gupta S, Mahan S, Sougnez C, Onofrio RC, Liefeld T, MacConaill L, Winckler W, Reich M, Li N, Mesirov JP, Gabriel SB, Getz G, Ardlie K, Chan V, Myer VE, Weber BL, Porter J, Warmuth M, Finan P, Harris JL, Meyerson M, Golub TR, Morrissey MP, Sellers WR, Schlegel R, Garraway LA. **The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity.** *Nature* 2012, **483**:603-607.
6. Barrett T, Troup DB, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Muerdtter RN, Holko M, Ayanbule O, Yefanov A, Soboleva A. **NCBI GEO: archive for functional genomics data sets--10 years on.** *Nucleic Acids Res* 2011, **39**:D1005-D1010.

7. Biankin AV, Waddell N, Kassahn KS, Gingras MC, Muthuswamy LB, Johns AL, Miller DK, Wilson PJ, Patch AM, Wu J, Chang DK, Cowley MJ, Gardiner BB, Song S, Harliwong I, Idrisoglu S, Nourse C, Nourbakhsh E, Manning S, Wani S, Gongora M, Pajic M, Scarlett CJ, Gill AJ, Pinho AV, Rooman I, Anderson M, Holmes O, Leonard C, Taylor D, Wood S, Xu Q, Nones K, Fink JL, Christ A, Bruxner T, Cloonan N, Kolle G, Newell F, Pinese M, Mead RS, Humphris JL, Kaplan W, Jones MD, Colvin EK, Nagrial AM, Humphrey ES, Chou A, Chin VT, Chantrill LA, Mawson A, Samra JS, Kench JG, Lovell JA, Daly RJ, Merrett ND, Toon C, Epari K, Nguyen NQ, Barbour A, Zeps N, Kakkar N, Zhao F, Wu YQ, Wang M, Muzny DM, Fisher WE, Brunicardi FC, Hodges SE, Reid JG, Drummond J, Chang K, Han Y, Lewis LR, Dinh H, Buhay CJ, Beck T, Timms L, Sam M, Begley K, Brown A, Pai D, Panchal A, Buchner N, De Borja R, Denroche RE, Yung CK, Serra S, Onetto N, Mukhopadhyay D, Tsao MS, Shaw PA, Petersen GM, Gallinger S, Hruban RH, Maitra A, Iacobuzio-Donahue CA, Schlick RD, Wolfgang CL, Morgan RA, Lawlor RT, Capelli P, Corbo V, Scardoni M, Tortora G, Tempero MA, Mann KM, Jenkins NA, Perez-Mancera PA, Adams DJ, Largaespada DA, Wessels LF, Rust AG, Stein LD, Tuveson DA, Copeland NG, Musgrove EA, Scarpa A, Eshleman JR, Hudson TJ, Sutherland RL, Wheeler DA, Pearson JV, McPherson JD, Gibbs RA, Grimmond SM. **Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes.** *Nature* 2012, **491**:399-405.
8. Biddle A, Mackenzie IC. **Cancer stem cells and EMT in carcinoma.** *Cancer and Metastasis Reviews* 2012, **31**:285-293.
9. Borgstrom P, Oh P, Czarny M, Racine B, Schnitzer JE. **Co-implanting orthotopic tissue creates stroma microenvironment enhancing growth and angiogenesis of multiple tumors.** *F1000Res* 2013, **2**:129.
10. Bradford JR, Farren M, Powell SJ, Runswick S, Weston SL, Brown H, Delpuech O, Wappett M, Smith NR, Carr TH, Dry JR, Gibson NJ, Barry ST. **RNA-Seq Differentiates Tumour and Host mRNA Expression Changes Induced by Treatment of Human Tumour Xenografts with the VEGFR Tyrosine Kinase Inhibitor Cediranib.** *PLoS One* 2013, **8**:e66003.
11. Bruce A. *Biología Molecular de la célula*. Omega. 2010.

12. Chaika NV, Yu F, Purohit V, Mehla K, Lazenby AJ, DiMaio D, Anderson JM, Yeh JJ, Johnson KR, Hollingsworth MA, Singh PK. **Differential expression of metabolic genes in tumor and stromal components of primary and metastatic loci in pancreatic adenocarcinoma.** *PLoS One* 2012, **7**:e32996.
13. Chang DK, Merrett ND, Biankin AV, NSW Pancreatic Cancer Network. **Improving outcomes for operable pancreatic cancer: is access to safer surgery the problem?** *J Gastroenterol Hepatol* 2008, **23**:1036-1045.
14. Chaudary N, Pintilie M, Schwock J, Dhani N, Clarke B, Milosevic M, Fyles A, Hill RP. **Characterization of the Tumor-Microenvironment in Patient-Derived Cervix Xenografts (OCICx).** *Cancers (Basel)* 2012, **4**:821-845.
15. Chiang DY, Villanueva A, Hoshida Y, Peix J, Newell P, Minguez B, LeBlanc AC, Donovan DJ, Thung SN, Solé M. **Focal gains of VEGFA and molecular classification of hepatocellular carcinoma.** *Cancer research* 2008, **68**:6779-6788.
16. Collisson EA, Sadanandam A, Olson P, Gibb WJ, Truitt M, Gu S, Cooc J, Weinkle J, Kim GE, Jakkula L, Feiler HS, Ko AH, Olshen AB, Danenberg KL, Tempero MA, Spellman PT, Hanahan D, Gray JW. **Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy.** *Nat Med* 2011, **17**:500-503.
17. Conway T, Wazny J, Bromage A, Tymms M, Sooraj D, Williams D, Beresford-Smith B. **Xenome--a tool for classifying reads from xenograft samples.** *Bioinformatics* 2012, **28**:i172-i178.
18. Cook N, Jodrell DI, Tuveson DA. **Predictive in vivo animal models and translation to clinical trials.** *Drug Discovery Today* 2012, **17**:253-260.
19. Cui H, Onyango P, Brandenburg S, Wu Y, Hsieh CL, Feinberg AP. **Loss of imprinting in colorectal cancer linked to hypomethylation of H19 and IGF2.** *Cancer Res* 2002, **62**:6442-6446.
20. Daniel VC, Marchionni L, Hierman JS, Rhodes JT, Devereux WL, Rudin CM, Yung R, Parmigiani G, Dorsch M, Peacock CD, Watkins DN. **A primary xenograft model of small-cell lung cancer reveals irreversible changes in gene expression imposed by culture in vitro.** *Cancer Res* 2009, **69**:3364-3373.

21. Davies K, Conlon KC. **Neuroendocrine tumors of the pancreas.** *Current gastroenterology reports* 2009, **11**:119-127.
22. DeRose YS, Wang G, Lin YC, Bernard PS, Buys SS, Ebbert MT, Factor R, Matsen C, Milash BA, Nelson E, Neumayer L, Randall RL, Stijleman IJ, Welm BE, Welm AL. **Tumor grafts derived from women with breast cancer authentically reflect tumor pathology, growth, metastasis and disease outcomes.** *Nat Med* 2011, **17**:1514-1520.
23. Edgar R, Domrachev M, Lash AE. **Gene Expression Omnibus: NCBI gene expression and hybridization array data repository.** *Nucleic acids research* 2002, **30**:207-210.
24. Eichhorn PJ, Rodón L, González-Juncà A, Dirac A, Gili M, Martínez-Sáez E, Aura C, Barba I, Peg V, Prat A, Cuartas I, Jimenez J, García-Dorado D, Sahuquillo J, Bernards R, Baselga J, Seoane J. **USP15 stabilizes TGF- β receptor I and promotes oncogenesis through the activation of TGF- β signaling in glioblastoma.** *Nat Med* 2012, **18**:429-435.
25. Ell B, Kang Y. **Transcriptional control of cancer metastasis.** *Trends Cell Biol* 2013, **23**:603-611.
26. Esteller M. **Cancer epigenomics: DNA methylomes and histone-modification maps.** *Nat Rev Genet* 2007, **8**:286-298.
27. Fichtner I, Rolff J, Soong R, Hoffmann J, Hammer S, Sommer A, Becker M, Merk J. **Establishment of patient-derived non-small cell lung cancer xenografts as models for the identification of predictive biomarkers.** *Clin Cancer Res* 2008, **14**:6456-6468.
28. Gadaleta E, Cutts RJ, Kelly GP, Crnogorac-Jurcevic T, Kocher HM, Lemoine NR, Chelala C. **A global insight into a cancer transcriptional space using pancreatic data: importance, findings and flaws.** *Nucleic Acids Research* 2011, **39**:7900-7907.
29. Gao L, He SB, Li DC. **Effects of miR-16 plus CA19-9 detections on pancreatic cancer diagnostic performance.** *Clin Lab* 2014, **60**:73-77.
30. Garrido-Laguna I, Uson M, Rajeshkumar NV, Tan AC, de Oliveira E, Karikari C, Villaroel MC, Salomon A, Taylor G, Sharma R, Hruban RH, Maitra A, Laheru D, Rubio-Viqueira B, Jimeno A, Hidalgo M. **Tumor engraftment in nude mice**

- and enrichment in stroma- related gene pathways predict poor survival and resistance to gemcitabine in patients with pancreatic cancer.** *Clin Cancer Res* 2011, **17**:5793-5800.
31. Ghaneh P, Costello E, Neoptolemos JP. **Biology and management of pancreatic cancer.** *Postgrad Med J* 2008, **84**:478-497.
32. Greenacre MJ, Blasius J. *Multiple correspondence analysis and related methods.* Chapman & Hall/CRC. 2006.
33. Hanahan D, Weinberg RA. **Hallmarks of cancer: the next generation.** *Cell* 2011, **144**:646-674.
34. Harada T, Chelala C, Bhakta V, Chaplin T, Caulee K, Baril P, Young BD, Lemoine NR. **Genome-wide DNA copy number analysis in pancreatic cancer using high-density single nucleotide polymorphism arrays.** *Oncogene* 2008, **27**:1951-1960.
35. Harada T, Chelala C, Crnogorac-Jurcevic T, Lemoine NR. **Genome-wide analysis of pancreatic cancer using microarray-based techniques.** *Pancreatology* 2009, **9**:13-24.
36. Harrell JC, Dye WW, Harvell DM, Sartorius CA, Horwitz KB. **Contaminating cells alter gene signatures in whole organ versus laser capture microdissected tumors: a comparison of experimental breast cancers and their lymph node metastases.** *Clin Exp Metastasis* 2008, **25**:81-88.
37. Herman JG, Baylin SB. **Gene silencing in cancer in association with promoter hypermethylation.** *New England Journal of Medicine* 2003, **349**:2042-2054.
38. Hidalgo M. **Pancreatic Cancer.** *The New England Journal of Medicine* 2010, **362**:1605-1617.
39. Hidalgo M. **New insights into pancreatic cancer biology.** *Annals of Oncology* 2012, **23**:x135-x138.
40. Hidalgo M, Amant F, Biankin V, Budinska E, Byrne T, Caldas C, Clarke B, de Jong S, Jonkers J, Maelandsmo M, Roman-Roman S, Seoane J, Trusolino L, Villanueva A, EurOPDX Consortium. **Patient-Derived Xenograft Models: An Emerging Platform for Translational Cancer Research.** *Cancer Discovery* 2014.

41. Hruban RH, Goggins M, Parsons J, Kern SE. **Progression model for pancreatic cancer.** *Clinical cancer research* 2000, **6**:2969-2972.
42. Hruban RH, Maitra A, Kern SE, Goggins M. **Precursors to pancreatic cancer.** *Gastroenterology Clinics of North America* 2007, **36**:831-849.
43. Hsu PP, Sabatini DM. **Cancer cell metabolism: Warburg and beyond.** *Cell* 2008, **134**:703-707.
44. Huynh H, Chow PKH, Palanisamy N, Salto-Tellez M, Goh BC, Lee CK, Somani A, Lee HS, Kalpana R, Yu K. **Bevacizumab and rapamycin induce growth suppression in mouse models of hepatocellular carcinoma.** *Journal of hepatology* 2008, **49**:52-60.
45. Huynh H, Soo KC, Chow PK, Panasci L, Tran E. **Xenografts of human hepatocellular carcinoma: a useful model for testing drugs.** *Clin Cancer Res* 2006, **12**:4306-4314.
46. Hwang RF, Moore T, Arumugam T, Ramachandran V, Amos KD, Rivera A, Ji B, Evans DB, Logsdon CD. **Cancer-associated stromal fibroblasts promote pancreatic tumor progression.** *Cancer Res* 2008, **68**:918-926.
47. Iacobuzio-Donahue CA, Klimstra DS, Adsay NV, Wilentz RE, Argani P, Sohn TA, Yeo CJ, Cameron JL, Kern SE, Hruban RH. **Dpc-4 protein is expressed in virtually all human intraductal papillary mucinous neoplasms of the pancreas: comparison with conventional ductal adenocarcinomas.** *Am J Pathol* 2000, **157**:755-761.
48. Iacobuzio-Donahue CA, Wilentz RE, Argani P, Yeo CJ, Cameron JL, Kern SE, Hruban RH. **Dpc4 protein in mucinous cystic neoplasms of the pancreas: frequent loss of expression in invasive carcinomas suggests a role in genetic progression.** *Am J Surg Pathol* 2000, **24**:1544-1548.
49. Jiao Y, Yonescu R, Offerhaus GJ, Klimstra DS, Maitra A, Eshleman JR, Herman JG, Poh W, Pelosof L, Wolfgang CL, Vogelstein B, Kinzler KW, Hruban RH, Papadopoulos N, Wood LD. **Whole-exome sequencing of pancreatic neoplasms with acinar differentiation.** *J Pathol* 2014, **232**:428-435.
50. Jimenez RE, Warshaw AL, Z'graggen K, Hartwig W, Taylor DZ, Compton CC, Fernández-del Castillo C. **Sequential accumulation of K-ras mutations and**

- p53 overexpression in the progression of pancreatic mucinous cystic neoplasms to malignancy.** *Ann Surg* 1999, **230**:501-511.
51. Jimeno A, Tan AC, Coffa J, Rajeshkumar NV, Kulesza P, Rubio-Viqueira B, Wheelhouse J, Diosdado B, Messersmith WA, Iacobuzio-Donahue C, Maitra A, Varella-Garcia M, Hirsch FR, Meijer GA, Hidalgo M. **Coordinated epidermal growth factor receptor pathway gene overexpression predicts epidermal growth factor receptor inhibitor sensitivity in pancreatic cancer.** *Cancer Res* 2008, **68**:2841-2849.
52. Jin K, Teng L, Shen Y, He K, Xu Z, Li G. **Patient-derived human tumour tissue xenografts in immunodeficient mice: a systematic review.** *Clin Transl Oncol* 2010, **12**:473-480.
53. Johnson JI, Decker S, Zaharevitz D, Rubinstein LV, Venditti JM, Schepartz S, Kalyandrug S, Christian M, Arbuck S, Hollingshead M, Sausville EA. **Relationships between drug activity in NCI preclinical in vitro and in vivo models and early clinical trials.** *Br J Cancer* 2001, **84**:1424-1431.
54. Jones PA, Baylin SB. **The fundamental role of epigenetic events in cancer.** *Nat Rev Genet* 2002, **3**:415-428.
55. Jones S, Zhang X, Parsons DW, Lin JC, Leary RJ, Angenendt P, Mankoo P, Carter H, Kamiyama H, Jimeno A, Hong SM, Fu B, Lin MT, Calhoun ES, Kamiyama M, Walter K, Nikolskaya T, Nikolsky Y, Hartigan J, Smith DR, Hidalgo M, Leach SD, Klein AP, Jaffee EM, Goggins M, Maitra A, Iacobuzio-Donahue C, Eshleman JR, Kern SE, Hruban RH, Karchin R, Papadopoulos N, Parmigiani G, Vogelstein B, Velculescu VE, Kinzler KW. **Core signaling pathways in human pancreatic cancers revealed by global genomic analyses.** *Science* 2008, **321**:1801-1806.
56. Jordan CT, Guzman ML, Noble M. **Cancer stem cells.** *New England Journal of Medicine* 2006, **355**:1253-1261.
57. Joyce JA, Pollard JW. **Microenvironmental regulation of metastasis.** *Nat Rev Cancer* 2009, **9**:239-252.
58. Kanda M, Matthaei H, Wu J, Hong SM, Yu J, Borges M, Hruban RH, Maitra A, Kinzler K, Vogelstein B, Goggins M. **Presence of somatic mutations in most**

- early-stage pancreatic intraepithelial neoplasia.** *Gastroenterology* 2012, **142**:730-733.e9.
59. Karhu R, Mahlamäki E, Kallioniemi A. **Pancreatic adenocarcinoma -- genetic portrait from chromosomes to microarrays.** *Genes Chromosomes Cancer* 2006, **45**:721-730.
 60. Kerbel RS, Guerin E, Francia G, Xu P, Lee CR, Ebos JM, Man S. **Preclinical recapitulation of antiangiogenic drug clinical efficacies using models of early or late stage breast cancer metastasis.** *Breast* 2013, **22**:S57-S65.
 61. Klöppel G, Solcia E, Longnecker DS, Capella C, Sobin LH. *Histological typing of tumours of the exocrine pancreas.* Springer. 1996.
 62. Kopetz S, Lemos R, Powis G. **The promise of patient-derived xenografts: the best laid plans of mice and men.** *Clin Cancer Res* 2012, **18**:5160-5162.
 63. Kroepil F, Fluegen G, Vallböhmer D, Baldus SE, Dizdar L, Raffel AM, Hafner D, Stoecklein NH, Knoefel WT. **Snail1 expression in colorectal cancer and its correlation with clinical and pathological parameters.** *BMC Cancer* 2013, **13**:145.
 64. Laurent C, Gentien D, Piperno-Neumann S, Némati F, Nicolas A, Tesson B, Desjardins L, Mariani P, Rapinat A, Sastre-Garau X, Couturier J, Hupé P, de Koning L, Dubois T, Roman-Roman S, Stern MH, Barillot E, Harbour JW, Saule S, Decaudin D. **Patient-derived xenografts recapitulate molecular features of human uveal melanomas.** *Mol Oncol* 2013, **7**:625-636.
 65. Li M, Marin-Muller C, Bharadwaj U, Chow KH, Yao Q, Chen C. **MicroRNAs: control and loss of control in human physiology and disease.** *World J Surg* 2009, **33**:667-684.
 66. Li F, Tiede B, Massagué J, Kang Y. **Beyond tumorigenesis: cancer stem cells in metastasis.** *Cell research* 2007, **17**:3-14.
 67. Liao YL, Sun YM, Chau GY, Chau YP, Lai TC, Wang JL, Horng JT, Hsiao M, Tsou AP. **Identification of SOX4 target genes using phylogenetic footprinting-based prediction from expression microarrays suggests that overexpression of SOX4 potentiates metastasis in hepatocellular carcinoma.** *Oncogene* 2008, **27**:5578-5589.

68. Lubezky N, Ben-Haim M, Lahat G, Marmor S, Solar I, Brazowski E, Nackache R, Klausner JM. **Intraductal papillary mucinous neoplasm of the pancreas: associated cancers, family history, genetic predisposition?** *Surgery* 2012, **151**:70-75.
69. Luebke AM, Baudis M, Matthaei H, Vashist YK, Verde PE, Hosch SB, Erbersdobler A, Klein CA, Izbicki JR, Knoefel WT, Stoecklein NH. **Losses at chromosome 4q are associated with poor survival in operable ductal pancreatic adenocarcinoma.** *Pancreatology* 2012, **12**:16-22.
70. Lukk M, Kapushesky M, Nikkilä J, Parkinson H, Goncalves A, Huber W, Ukkonen E, Brazma A. **A global map of human gene expression.** *Nature biotechnology* 2010, **28**:322-324.
71. Lum DH, Matsen C, Welm AL, Welm BE. **Overview of human primary tumorgraft models: comparisons with traditional oncology preclinical models and the clinical relevance and utility of primary tumorgrafts in basic and translational oncology research.** *Curr Protoc Pharmacol* 2012, **Chapter 14**:Unit 14.22.
72. Macian F. **NFAT proteins: key regulators of T-cell development and function.** *Nature Reviews Immunology* 2005, **5**:472-484.
73. Maitra A, Hruban RH. **Pancreatic cancer.** *Annu Rev Pathol* 2008, **3**:157-188.
74. Mani SA, Guo W, Liao M-J, Eaton EN, Ayyanan A, Zhou AY, Brooks M, Reinhard F, Zhang CC, Shipitsin M. **The epithelial-mesenchymal transition generates cells with properties of stem cells.** *Cell* 2008, **133**:704-715.
75. Marangoni E, Vincent-Salomon A, Auger N, Degeorges A, Assayag F, de Cremoux P, de Plater L, Guyader C, De Pinieux G, Judde JG, Rebucci M, Tran-Perennou C, Sastre-Garau X, Sigal-Zafrani B, Delattre O, Diéras V, Poupon MF. **A new model of patient tumor-derived breast cancer xenografts for preclinical assays.** *Clin Cancer Res* 2007, **13**:3989-3998.
76. Martinez-Garcia R, Juan D, Rausell A, Muñoz M, Baños N, Menéndez C, Lopez-Casas PP, Rico D, Valencia A, Hidalgo M. **Transcriptional dissection of pancreatic tumors engrafted in mice.** *Genome Med* 2014, **6**:27.

77. Martínez-Trillos A, Quesada V, Villamor N, Puente XS, López-Otín C, Campo E. **Recurrent Gene Mutations in CLL.** In *Advances in Chronic Lymphocytic Leukemia*. Springer. 2013. 87-107.
78. Matthews L, Gopinath G, Gillespie M, Caudy M, Croft D, de Bono B, Garapati P, Hemish J, Hermjakob H, Jassal B, Kanapin A, Lewis S, Mahajan S, May B, Schmidt E, Vastrik I, Wu G, Birney E, Stein L, D'Eustachio P. **Reactome knowledgebase of human biological pathways and processes.** *Nucleic Acids Res* 2009, **37**:D619-D622.
79. Mattie M, Christensen A, Chang MS, Yeh W, Said S, Shostak Y, Capo L, Verlinsky A, An Z, Joseph I, Zhang Y, Kumar-Ganesan S, Morrison K, Stover D, Challita-Eid P. **Molecular characterization of patient-derived human pancreatic tumor xenograft models for preclinical and translational development of cancer therapeutics.** *Neoplasia* 2013, **15**:1138-1150.
80. Maupin KA, Sinha A, Eugster E, Miller J, Ross J, Paulino V, Keshamouni VG, Tran N, Berens M, Webb C, Haab BB. **Glycogene expression alterations associated with pancreatic cancer epithelial-mesenchymal transition in complementary model systems.** *PLoS One* 2010, **5**:e13002.
81. McCall MN, Bolstad BM, Irizarry RA. **Frozen robust multiarray analysis (fRMA).** *Biostatistics* 2010, **11**:242-253.
82. McCall MN, Uppal K, Jaffee HA, Zilliox MJ, Irizarry RA. **The Gene Expression Barcode: leveraging public data repositories to begin cataloging the human and murine transcriptomes.** *Nucleic Acids Res* 2011, **39**:D1011-D1015.
83. Meric-Bernstam F, Mills GB. **Overcoming implementation challenges of personalized cancer therapy.** *Nat Rev Clin Oncol* 2012, **9**:542-548.
84. Monsma DJ, Monks NR, Cherba DM, Dylewski D, Eugster E, Jahn H, Srikanth S, Scott SB, Richardson PJ, Everts RE, Ishkin A, Nikolsky Y, Resau JH, Sigler R, Nickoloff BJ, Webb CP. **Genomic characterization of explant tumorgraft models derived from fresh patient tumor tissue.** *J Transl Med* 2012, **10**:125.
85. Murphy SJ, Hart SN, Lima JF, Kipp BR, Klebig M, Winters JL, Szabo C, Zhang L, Eckloff BW, Petersen GM, Scherer SE, Gibbs RA, McWilliams RR, Vasmatzis G, Couch FJ. **Genetic alterations associated with progression from**

- pancreatic intraepithelial neoplasia to invasive pancreatic tumor.** *Gastroenterology* 2013, **145**:1098-1109.e1.
86. Nguyen DX, Bos PD, Massagué J. **Metastasis: from dissemination to organ-specific colonization.** *Nat Rev Cancer* 2009, **9**:274-284.
87. Nones K, Waddell N, Song S, Patch AM, Miller D, Johns A, Wu J, Kassahn KS, Wood D, Bailey P, Fink L, Manning S, Christ AN, Nourse C, Kazakoff S, Taylor D, Leonard C, Chang DK, Jones MD, Thomas M, Watson C, Pinese M, Cowley M, Rooman I, Pajic M, Butturini G, Malpaga A, Corbo V, Crippa S, Falconi M, Zamboni G, Castelli P, Lawlor RT, Gill AJ, Scarpa A, Pearson JV, Biankin AV, Grimmond SM. **Genome-wide DNA methylation patterns in pancreatic ductal adenocarcinoma reveal epigenetic deregulation of SLIT-ROBO, ITGA2 and MET signaling.** *Int J Cancer* 2014, **135**:1110-1118.
88. Park B, Jeong BC, Choi YL, Kwon GY, Lim JE, Seo SI, Jeon SS, Lee HM, Choi HY, Lee KS. **Development and characterization of a bladder cancer xenograft model using patient-derived tumor tissue.** *Cancer Sci* 2013, **104**:631-638.
89. Pàez-Ribes M, Allen E, Hudock J, Takeda T, Okuyama H, Viñals F, Inoue M, Bergers G, Hanahan D, Casanovas O. **Antiangiogenic therapy elicits malignant progression of tumors to increased local invasion and distant metastasis.** *Cancer Cell* 2009, **15**:220-231.
90. Pei H, Li L, Fridley BL, Jenkins GD, Kalari KR, Lingle W, Petersen G, Lou Z, Wang L. **FKBP51 affects cancer cell response to chemotherapy by negatively regulating Akt.** *Cancer Cell* 2009, **16**:259-266.
91. Puente XS, Pinyol M, Quesada V, Conde L, Ordóñez GR, Villamor N, Escaramis G, Jares P, Beà S, González-Díaz M, Bassaganyas L, Baumann T, Juan M, López-Guerra M, Colomer D, Tubío JM, López C, Navarro A, Tornador C, Aymerich M, Rozman M, Hernández JM, Puente DA, Freije JM, Velasco G, Gutiérrez-Fernández A, Costa D, Carrió A, Guijarro S, Enjuanes A, Hernández L, Yagüe J, Nicolás P, Romeo-Casabona CM, Himmelbauer H, Castillo E, Dohm JC, de Sanjosé S, Piris MA, de Alava E, San Miguel J, Royo R, Gelpí JL, Torrents D, Orozco M, Pisano DG, Valencia A, Guigó R, Bayés M, Heath S, Gut M, Klatt P, Marshall J, Raine K, Stebbings LA, Futreal PA, Stratton MR,

- Campbell PJ, Gut I, López-Guillermo A, Estivill X, Montserrat E, López-Otín C, Campo E. **Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia.** *Nature* 2011, **475**:101-105.
92. Quesada V, Ramsay AJ, Rodríguez D, Puente XS, Campo E, López-Otín C. **The genomic landscape of chronic lymphocytic leukemia: clinical implications.** *BMC Med* 2013, **11**:124.
93. Rausell A, Juan D, Pazos F, Valencia A. **Protein interactions and ligand binding: from protein subfamilies to functional specificity.** *Proc Natl Acad Sci U S A* 2010, **107**:1995-2000.
94. Reid MD, Saka B, Balci S, Goldblum AS, Adsay NV. **Molecular genetics of pancreatic neoplasms and their morphologic correlates: an update on recent advances and potential diagnostic applications.** *Am J Clin Pathol* 2014, **141**:168-180.
95. Reyat F, Guyader C, Decraene C, Lucchesi C, Auger N, Assayag F, De Plater L, Gentien D, Poupon MF, Cottu P, De Cremoux P, Gestraud P, Vincent-Salomon A, Fontaine JJ, Roman-Roman S, Delattre O, Decaudin D, Marangoni E. **Molecular profiling of patient-derived breast cancer xenografts.** *Breast Cancer Res* 2012, **14**:R11.
96. Rosa-Ribeiro R, Nishan U, Vidal RO, Barbosa GO, Reis LO, Cesar CL, Carvalho HF. **Transcription factors involved in prostate gland adaptation to androgen deprivation.** *PLoS One* 2014, **9**:e97080.
97. Rubio-Viqueira B, Hidalgo M. **Direct in vivo xenograft tumor model for predicting chemotherapeutic drug response in cancer patients.** *Clin Pharmacol Ther* 2009, **85**:217-221.
98. Rubio-Viqueira B, Jimeno A, Cusatis G, Zhang X, Iacobuzio-Donahue C, Karikari C, Shi C, Danenberg K, Danenberg PV, Kuramochi H, Tanaka K, Singh S, Salimi-Moosavi H, Bouraoud N, Amador ML, Altiock S, Kulesza P, Yeo C, Messersmith W, Eshleman J, Hruban RH, Maitra A, Hidalgo M. **An in vivo platform for translational drug development in pancreatic cancer.** *Clin Cancer Res* 2006, **12**:4652-4661.

99. Ruggeri BA, Camp F, Miknyoczki S. **Animal models of disease: pre-clinical animal models of cancer and their applications and utility in drug discovery.** *Biochem Pharmacol* 2014, **87**:150-161.
100. Samuels AL, Peeva VK, Papa RA, Firth MJ, Francis RW, Beesley AH, Lock RB, Kees UR. **Validation of a mouse xenograft model system for gene expression analysis of human acute lymphoblastic leukaemia.** *BMC Genomics* 2010, **11**:256.
101. Sausville EA, Burger AM. **Contributions of human tumor xenografts to anticancer drug development.** *Cancer Res* 2006, **66**:3351-3354.
102. Schoedel KE, Finkelstein SD, Ohori NP. **K-Ras and microsatellite marker analysis of fine-needle aspirates from intraductal papillary mucinous neoplasms of the pancreas.** *Diagn Cytopathol* 2006, **34**:605-608.
103. Sivanand S, Peña-Llopis S, Zhao H, Kucejova B, Spence P, Pavia-Jimenez A, Yamasaki T, McBride DJ, Gillen J, Wolff NC, Morlock L, Lotan Y, Raj GV, Sagalowsky A, Margulis V, Cadeddu JA, Ross MT, Bentley DR, Kabbani W, Xie XJ, Kapur P, Williams NS, Brugarolas J. **A validated tumorgraft model reveals activity of dovitinib against renal cell carcinoma.** *Sci Transl Med* 2012, **4**:137ra75.
104. Smyth GK, Michaud J, Scott HS. **Use of within-array replicate spots for assessing differential expression in microarray experiments.** *Bioinformatics* 2005, **21**:2067-2075.
105. Sobin LH, Gospodarowicz MK, Wittekind C. *TNM classification of malignant tumours.* Wiley. 2011.
106. Sparrow S, Jones M, Billington S, Stace B. **The in vivo malignant transformation of mouse fibroblasts in the presence of human tumour xenografts.** *Br J Cancer* 1986, **53**:793-797.
107. Srivastava SK, Arora S, Singh S, Bhardwaj A, Averett C, Singh AP. **MicroRNAs in pancreatic malignancy: Progress and promises.** *Cancer Lett* 2014, **347**:167-174.
108. Steeg PS. **Metastasis suppressors alter the signal transduction of cancer cells.** *Nat Rev Cancer* 2003, **3**:55-63.

109. Stratford JK, Bentrem DJ, Anderson JM, Fan C, Volmar KA, Marron JS, Routh ED, Caskey LS, Samuel JC, Der CJ, Thorne LB, Calvo BF, Kim HJ, Talamonti MS, Iacobuzio-Donahue CA, Hollingsworth MA, Perou CM, Yeh JJ. **A six-gene signature predicts survival of patients with localized pancreatic ductal adenocarcinoma.** *PLoS Med* 2010, **7**:e1000307.
110. Subhi AL, Tang B, Balsara BR, Altomare DA, Testa JR, Cooper HS, Hoffman JP, Meropol NJ, Kruger WD. **Loss of methylthioadenosine phosphorylase and elevated ornithine decarboxylase is common in pancreatic cancer.** *Clin Cancer Res* 2004, **10**:7290-7296.
111. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. **Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.** *Proc Natl Acad Sci U S A* 2005, **102**:15545-15550.
112. Suzuki H, Forrest AR, van Nimwegen E, Daub CO, Balwierz PJ, Irvine KM, Lassmann T, Ravasi T, Hasegawa Y, de Hoon MJ, Katayama S, Schroder K, Carninci P, Tomaru Y, Kanamori-Katayama M, Kubosaki A, Akalin A, Ando Y, Arner E, Asada M, Asahara H, Bailey T, Bajic VB, Bauer D, Beckhouse AG, Bertin N, Björkegren J, Brombacher F, Bulger E, Chalk AM, Chiba J, Cloonan N, Dawe A, Dostie J, Engström PG, Essack M, Faulkner GJ, Fink JL, Fredman D, Fujimori K, Furuno M, Gojobori T, Gough J, Grimmond SM, Gustafsson M, Hashimoto M, Hashimoto T, Hatakeyama M, Heinzl S, Hide W, Hofmann O, Hörnquist M, Huminiecki L, Ikeo K, Imamoto N, Inoue S, Inoue Y, Ishihara R, Iwayanagi T, Jacobsen A, Kaur M, Kawaji H, Kerr MC, Kimura R, Kimura S, Kimura Y, Kitano H, Koga H, Kojima T, Kondo S, Konno T, Krogh A, Kruger A, Kumar A, Lenhard B, Lennartsson A, Lindow M, Lizio M, Macpherson C, Maeda N, Maher CA, Maqungo M, Mar J, Matigian NA, Matsuda H, Mattick JS, Meier S, Miyamoto S, Miyamoto-Sato E, Nakabayashi K, Nakachi Y, Nakano M, Nygaard S, Okayama T, Okazaki Y, Okuda-Yabukami H, Orlando V, Otomo J, Pachkov M, Petrovsky N, Plessy C, Quackenbush J, Radovanovic A, Rehli M, Saito R, Sandelin A, Schmeier S, Schönbach C, Schwartz AS, Semple CA, Sera M, Severin J, Shirahige K, Simons C, St Laurent G, Suzuki

- M, Suzuki T, Sweet MJ, Taft RJ, Takeda S, Takenaka Y, Tan K, Taylor MS, Teasdale RD, Tegnér J, Teichmann S, Valen E, Wahlestedt C, Waki K, Waterhouse A, Wells CA, Winther O, Wu L, Yamaguchi K, Yanagawa H, Yasuda J, Zavolan M, Hume DA, Arakawa T, Fukuda S, Imamura K, Kai C, Kaiho A, Kawashima T, Kawazu C, Kitazume Y, Kojima M, Miura H, Murakami K, Murata M, Ninomiya N, Nishiyori H, Noma S, Ogawa C, Sano T, Simon C, Tagami M, Takahashi Y, Kawai J, Hayashizaki Y. **The transcriptional network that controls growth arrest and differentiation in a human myeloid leukemia cell line.** *Nat Genet* 2009, **41**:553-562.
113. Tentler JJ, Tan AC, Weekes CD, Jimeno A, Leong S, Pitts TM, Arcaroli JJ, Messersmith WA, Eckhardt SG. **Patient-derived tumour xenografts as models for oncology drug development.** *Nat Rev Clin Oncol* 2012, **9**:338-350.
 114. Uronis JM, Osada T, McCall S, Yang XY, Mantyh C, Morse MA, Lyerly HK, Clary BM, Hsu DS. **Histological and molecular evaluation of patient-derived colorectal cancer explants.** *PLoS One* 2012, **7**:e38422.
 115. Valastyan S, Weinberg RA. **Tumor metastasis: molecular insights and evolving paradigms.** *Cell* 2011, **147**:275-292.
 116. Valdes C, Seo P, Tsinoremas N, Clarke J. **Characteristics of cross-hybridization and cross-alignment of expression in pseudo-xenograft samples by RNA-Seq and microarrays.** *J Clin Bioinforma* 2013, **3**:8.
 117. Van den Broeck A, Vankelecom H, Van Eijsden R, Govaere O, Topal B. **Molecular markers associated with outcome and metastasis in human pancreatic cancer.** *J Exp Clin Cancer Res* 2012, **31**:68.
 118. Vanharanta S, Massagué J. **Origins of metastatic traits.** *Cancer Cell* 2013, **24**:410-421.
 119. Villarroel MC, Rajeshkumar NV, Garrido-Laguna I, De Jesus-Acosta A, Jones S, Maitra A, Hruban RH, Eshleman JR, Klein A, Laheru D, Donehower R, Hidalgo M. **Personalizing cancer treatment in the age of global genomic analyses: PALB2 gene mutations and the response to DNA damaging agents in pancreatic cancer.** *Mol Cancer Ther* 2011, **10**:3-8.
 120. Walters DM, Stokes JB, Adair SJ, Stelow EB, Borgman CA, Lowrey BT, Xin W, Blais EM, Lee JK, Papin JA, Parsons JT, Bauer TW. **Clinical, molecular**

- and genetic validation of a murine orthotopic xenograft model of pancreatic adenocarcinoma using fresh human specimens.** *PLoS One* 2013, **8**:e77065.
121. Warburg O. **On the origin of cancer cells.** *Science* 1956, **123**:309-314.
 122. Wartha K, Herting F, Hasmann M. **Fit-for purpose use of mouse models to improve predictivity of cancer therapeutics evaluation.** *Pharmacol Ther* 2014, **142**:351-361.
 123. Wood WM, Etemad S, Yamamoto M, Goldhamer DJ. **MyoD-expressing progenitors are essential for skeletal myogenesis and satellite cell development.** *Dev Biol* 2013, **384**:114-127.
 124. Yachida S, Jones S, Bozic I, Antal T, Leary R, Fu B, Kamiyama M, Hruban RH, Eshleman JR, Nowak MA, Velculescu VE, Kinzler KW, Vogelstein B, Iacobuzio-Donahue CA. **Distant metastasis occurs late during the genetic evolution of pancreatic cancer.** *Nature* 2010, **467**:1114-1117.
 125. Yeh JE, Toniolo PA, Frank DA. **Targeting transcription factors: promising new strategies for cancer therapy.** *Curr Opin Oncol* 2013, **25**:652-658.
 126. Yonezawa S, Higashi M, Yamada N, Goto M. **Precursor lesions of pancreatic cancer.** *Gut Liver* 2008, **2**:137-154.
 127. Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, Treviño V, Shen H, Laird PW, Levine DA, Carter SL, Getz G, Stemke-Hale K, Mills GB, Verhaak RG. **Inferring tumour purity and stromal and immune cell admixture from expression data.** *Nat Commun* 2013, **4**:2612.
 128. Z'graggen K, Rivera JA, Compton CC, Pins M, Werner J, Fernández-del Castillo C, Rattner DW, Lewandrowski KB, Rustgi AK, Warshaw AL. **Prevalence of activating K-ras mutations in the evolutionary stages of neoplasia in intraductal papillary mucinous tumors of the pancreas.** *Ann Surg* 1997, **226**:491-500.

ANEXO I – Material suplementario

Tabla S1. Datos de todos los perfiles de expresión utilizados en este trabajo, incluyendo aquellos análisis de los que han formado parte. Los datos procedentes del *CCLE Portal*, se han obtenido a través de su página web, <http://www.broadinstitute.org/ccle>.

Nombre muestra	Tipo muestra – Tejido de origen	Procedencia	Publicación	Resultados
GSM388115, GSM388117, GSM388119, GSM388121-GSM388153	Tumor primario completo – Páncreas	GSE15471	Badea et al. 2008	Figuras 7-8, 10-14, 16-30
GSM414927, GSM414929, GSM414933, GSM414937, GSM414939, GSM414941, GSM414946, GSM414952, GSM414954, GSM414956, GSM414962, GSM414965, GSM414969, GSM414971, GSM414974	Tumor primario completo – Páncreas	GSE16515	Pei et al. 2009	Figuras 7-8, 10-14, 16-30
GSM1252927, GSM1252928	Tumor primario completo – Páncreas	GSE51798	Martinez-Garcia et al. 2014	Figuras 7-8, 10-30, 34-37, 39
TP_Panc033	Tumor primario completo – Páncreas	CNIO	No publicado	Figuras 34-37
TP_Panc281, TP_Panc287, TP_JH033	Tumor primario completo – Páncreas	JH	Garrido-Laguna et al. 2011	Figuras 31-33
GSM463731, GSM463735, GSM463739, GSM463743, GSM1053826, GSM1053829, GSM1053830, GSM1053833, GSM1053834, GSM1053837, GSM1053838, GSM1053841, GSM1053843, GSM1053845, GSM1053846	Tumor primario completo – Páncreas	GSE42952	Van den Broeck et al. 2012	Figuras 34-37 Tabla 5

GSM536033, GSM831958-GSM831960, GSM831964, GSM831966, GSM831970, GSM831973, GSM831988, GSM831992, GSM832006-GSM832007, GSM832010-GSM832012	Tumor primario completo – Páncreas	GSE34153	Chaika et al. 2012	Tabla 5
GSM446731-GSM446757	Tumor primario microdisecionado – Páncreas	GSE17891	Collisson et al. 2011	Figuras 22-28
Met_Panc037, Met_Panc042Hep, Met_Panc042Peri, Met_Panc037Hep, Met_Panc016, Met_Z11.08, Met_Z12.419.A2, Met_Z12.437	Metástasis fresca humana – Páncreas	CNIO	No publicado	Figuras 34-37, 39
GSM1053825, GSM1053827, GSM1053828, GSM1053831, GSM1053832, GSM1053835, GSM1053836, GSM1053839, GSM1053840, GSM1053842, GSM1053844, GSM1053847	Metástasis fresca humana – Páncreas	GSE42952	Van den Broeck et al. 2012	Figuras 34-37 Tabla 5
GSM831943-GSM831957, GSM831961-GSM831963, GSM831965, GSM831967-GSM831969, GSM831971-GSM831972, GSM831974-GSM831987, GSM831989-GSM831991, GSM831993-GSM832005, GSM832008-GSM832009, GSM832013-GSM832016	Metástasis fresca humana – Páncreas	GSE34153	Chaika et al. 2012	Tabla 5
GSM1252892-GSM1252926	Modelo PDX – Páncreas	GSE51798	Martinez-Garcia et al. 2014	Figuras 7-8, 10-12, 16-30, 34-37
PDX_Panc042, PDX_Panc047	Modelo PDX – Páncreas	CNIO	No publicado	Figuras 15, 34-39 Tabla 6

GSM242823, GSM242825, GSM242827, GSM242831, GSM242833, GSM242835, GSM242837, GSM242839, GSM242841, GSM242843, GSM242845, GSM242847, GSM242849, GSM242851, GSM242853, GSM242855, GSM242857	Modelo PDX – Páncreas	GSE9599	Jimeno et al. 2008	Figuras 7-8, 10-12, 16-30
Panc281_F5_1, Panc281_F5_2, Panc281_F10_1, Panc281_F10_2, Panc286_F5_1, Panc286_F5_2, Panc286_F10_1, Panc286_F10_2, Panc287_F5_1, Panc287_F5_2, Panc287_F10_1, Panc287_F10_2, JH033_F5_2, JH033_F10_1, JH033_F10_2	Modelo PDX – Páncreas	JH	Garrido-Laguna et al. 2011	Figuras 31-33
GSM1129552, GSM1129554, GSM1129555, GSM1129564, GSM1129566, GSM1129568, GSM1129569, GSM1129573, GSM1129575, GSM1129576, GSM1129583, GSM1129584, GSM1129585, GSM1129588, GSM1129593	Modelo PDX ortotópico – Páncreas	GSE46385	Walters et al. 2013	Figura 8
PDXMet_A6L, PDXMet_Panc010, PDXMet_Panc011, PDXMet_Panc016, PDXMet_Panc019, PDXMet_Panc020, PDXMet_Panc025, PDXMet_Panc042, PDXMet_Panc047	Modelo PDX derivado de metástasis – Páncreas	CNIO	No publicado	Figuras 34-39 Tabla 6
GSM446765-GSM446782	Línea celular – Páncreas	GSE17891	Collisson et al. 2011	Figuras 7-8, 13-14, 16-20
GSM540374-GSM540395	Línea celular – Páncreas	GSE21654	Maupin et al. 2010	Figuras 7-8, 13-14, 16-20
GSM139131, GSM143545-GSM143553	Tumor primario completo – Hígado	GSE6222	Liao et al. 2008	Figuras 9-14, 16-21

GSM148605-GSM148611	Tumor primario completo – Hígado	GSE6465	Huynh et al. 2008	Figuras 9-14, 16-21
GSM248688-GSM248778	Tumor primario completo – Hígado	GSE9843	Chiang et al. 2008	Figuras 9-14, 16-21
GSM148598-GSM148646	Modelo PDX – Hígado	GSE6465	Huynh et al. 2008	Figuras 9-12, 16-21
240866, 241034, 241064, 241090, 241102, 241120, 257660, 383606, 383636, 383642, 383654, 383774, 419730, 436462, 436554, 436592, 454756, 569420, 569542, 587608, 587624, 587650, 587704, 587732, 698850, 698862, 698870, 779652	Línea celular – Hígado	<i>CCL</i> E Portal	Barretina et al. 2012	Figuras 9, 13-14, 16-20
GSM138620	Línea celular – Hígado	GSE6222	Liao et al. 2008	Figuras 9, 13-14, 16-20
GSM148594-GSM148597	Línea celular – Hígado	GSE6465	Huynh et al. 2008	Figuras 9, 13-14, 16-20

Tabla S2. Datos de los genes más significativamente enriquecidos (GSEA *core genes*) en el análisis de enriquecimiento funcional. Para cada gen se muestra la frecuencia de expresión en los tumores primarios y en los modelos PDX, así como el grupo funcional en el que se encuentra enriquecido.

Gen	Frecuencia de expresión en modelos PDX	Frecuencia de expresión en tumores primarios	Ciclo celular	Replicación del ADN	Transducción de señales	Hemostasia	Organización de la matriz extracelular
A2M	0	1	No	No	No	Si	No
ACTR1A	0.17	0	Si	No	No	No	No
ADAMTS2	0.02	0.98	No	No	No	No	Si
ADCY7	0.38	0.75	No	No	Si	No	No
ADCY9	0.17	0.68	No	No	Si	No	No
ADCYAP1	0	0.28	No	No	Si	No	No
ADORA3	0	0.13	No	No	Si	No	No
ADRA2A	0.72	0.85	No	No	Si	No	No
AGT	0.08	0.74	No	No	Si	No	No
AGTR1	0.17	0.36	No	No	Si	No	No
ALB	0	0.47	No	No	No	Si	No
APITD1	0.43	0	Si	Si	No	No	No
APLNR	0	0.17	No	No	Si	No	No
ATP2B4	0.70	1	No	No	No	Si	No
AURKA	1	0.19	Si	No	No	No	No
AURKB	0.49	0	Si	Si	No	No	No
BIRC5	0.94	0.08	Si	Si	No	No	No
BRCA1	0.19	0	Si	No	No	No	No
BUB1	0.87	0.13	Si	Si	No	No	No
BUB1B	0.98	0.19	Si	Si	No	No	No
C1QA	0	1	No	No	Si	No	No
C1QB	0	1	No	No	Si	No	No
C1QC	0	1	No	No	Si	No	No
C1R	0.04	1	No	No	Si	No	No
C1S	0.08	1	No	No	Si	No	No
C3	0.21	1	No	No	Si	No	No
C3AR1	0	0.68	No	No	Si	No	No
C5	0	0.06	No	No	Si	No	No
C5AR1	0	0.85	No	No	Si	No	No
CALCRL	0	0.42	No	No	Si	No	No

CASC5	0.87	0.08	Si	Si	No	No	No
CCL11	0	0.21	No	No	Si	No	No
CCL19	0	0.81	No	No	Si	No	No
CCL2	0	0.98	No	No	Si	No	No
CCL20	0.68	0.79	No	No	Si	No	No
CCL21	0	0.30	No	No	Si	No	No
CCL4	0	0.43	No	No	Si	No	No
CCL5	0.19	0.83	No	No	Si	No	No
CCNA2	0.79	0.06	Si	Si	No	No	No
CCNB1	0.94	0.34	Si	No	No	No	No
CCNB2	0.98	0.28	Si	No	No	No	No
CCNE2	0.55	0.02	Si	No	No	No	No
CCR1	0	0.45	No	No	Si	No	No
CCR2	0	0.32	No	No	Si	No	No
CCR5	0	0.21	No	No	Si	No	No
CCR7	0	0.23	No	No	Si	No	No
CD36	0.04	0.75	No	No	No	Si	No
CDC20	0.96	0.19	Si	Si	No	No	No
CDC23	0.36	0.11	Si	No	No	No	No
CDC25B	0.85	0.34	Si	No	No	No	No
CDC25C	0.38	0	Si	No	No	No	No
CDC45	0.28	0	Si	No	No	No	No
CDC6	0.72	0.08	Si	Si	No	No	No
CDC7	0.17	0.02	Si	Si	No	No	No
CDCA8	0.64	0	Si	Si	No	No	No
CDK1	1	0.64	Si	No	No	No	No
CDK2	0.75	0.06	Si	Si	No	No	No
CDKN2A	0.28	0.13	Si	No	No	No	No
CDKN2C	0.49	0.11	Si	No	No	No	No
CDT1	0.30	0	Si	Si	No	No	No
CENPA	0.89	0.11	Si	Si	No	No	No
CENPE	0.55	0	Si	Si	No	No	No
CENPF	0.87	0.15	Si	Si	No	No	No
CENPH	0.36	0	Si	Si	No	No	No
CENPK	0.87	0.32	Si	Si	No	No	No
CENPM	0.21	0	Si	Si	No	No	No
CENPN	0.45	0	Si	Si	No	No	No
CFD	0.15	0.77	No	No	Si	Si	No
CHEK1	0.38	0.09	Si	No	No	No	No
CHRM3	0.08	0.17	No	No	Si	No	No
CLU	0.43	1	No	No	No	Si	No

CNR1	0	0.08	No	No	Si	No	No
COL10A1	0	0.98	No	No	No	No	Si
COL11A1	0	0.85	No	No	No	No	Si
COL12A1	0.30	1	No	No	No	No	Si
COL14A1	0	0.92	No	No	No	No	Si
COL15A1	0.04	1	No	No	No	No	Si
COL16A1	0.08	0.85	No	No	No	No	Si
COL18A1	0.11	0.85	No	No	No	No	Si
COL27A1	0.47	0.91	No	No	No	No	Si
COL4A1	0.58	1	No	No	Si	No	Si
COL4A2	0.47	1	No	No	Si	No	Si
COL4A5	0.26	0.70	No	No	Si	No	Si
COL5A1	0.04	1	No	No	No	No	Si
COL5A2	0.28	1	No	No	No	No	Si
COL6A1	0.53	1	No	No	No	No	Si
COL6A2	0.08	1	No	No	No	No	Si
COL6A3	0	1	No	No	No	No	Si
COL8A1	0.02	1	No	No	No	No	Si
COL8A2	0.04	0.62	No	No	No	No	Si
CRH	0	0.11	No	No	Si	No	No
CSNK1D	0.87	0.49	Si	No	No	No	No
CSNK1E	0.32	0.02	Si	No	No	No	No
CTRB2	0	0.72	No	No	No	No	Si
CX3CL1	0.21	0.28	No	No	Si	No	No
CX3CR1	0	0.70	No	No	Si	No	No
CXCL10	0.08	0.53	No	No	Si	No	No
CXCL11	0.19	0.32	No	No	Si	No	No
CXCL12	0.06	1	No	No	Si	No	No
CXCL13	0	0.53	No	No	Si	No	No
CXCL2	0.77	0.98	No	No	Si	No	No
CXCL6	0.28	0.77	No	No	Si	No	No
CXCL9	0	0.47	No	No	Si	No	No
CXCR2	0	0.06	No	No	Si	No	No
CXCR4	0.28	1	No	No	Si	No	No
CXCR7	0.04	0.96	No	No	Si	No	No
CYSLTR1	0	0.15	No	No	Si	No	No
DARC	0	0.40	No	No	Si	No	No
DBF4	0.72	0.09	Si	Si	No	No	No
DCTN1	0.98	0.51	Si	No	No	No	No
DHFR	0.87	0.11	Si	No	No	No	No
DIDO1	0.28	0.02	Si	No	No	No	No

E2F4	0.96	0.36	Si	No	No	No	No
E2F5	0.23	0	Si	No	No	No	No
EDN1	0.75	0.89	No	No	Si	No	No
EDN3	0.13	0.25	No	No	Si	No	No
EDNRA	0	1	No	No	Si	No	No
EDNRB	0.06	0.81	No	No	Si	No	No
EGF	0.02	0.13	No	No	No	Si	No
EMR2	0.02	0.17	No	No	Si	No	No
F13A1	0	0.98	No	No	No	Si	No
F2R	0.09	1	No	No	Si	No	No
F2RL2	0.04	0.75	No	No	Si	No	No
F8	0.02	0.13	No	No	No	Si	No
FBXO5	0.19	0.02	Si	Si	No	No	No
FEN1	0.36	0	Si	Si	No	No	No
FGA	0	0.11	No	No	No	Si	No
FGB	0	0.11	No	No	No	Si	No
FGG	0.02	0.36	No	No	No	Si	No
FGR	0	0.09	No	No	No	Si	No
FLNA	0.74	1	No	No	No	Si	No
FN1	0.42	1	No	No	Si	Si	No
FPR1	0	0.51	No	No	Si	No	No
FPR3	0	0.81	No	No	Si	No	No
FZD1	0.11	0.94	No	No	Si	No	No
FZD7	0.57	0.98	No	No	Si	No	No
GAS6	0.75	1	No	No	No	Si	No
GCG	0	0.83	No	No	Si	No	No
GINS1	0.96	0.21	Si	Si	No	No	No
GINS2	0.70	0.04	Si	Si	No	No	No
GNAI1	0.64	1	No	No	Si	No	No
GNAI2	0.55	0.77	No	No	Si	No	No
GNAL	0	0.08	No	No	Si	No	No
GNG11	0.08	0.98	No	No	Si	Si	No
GNG2	0.04	1	No	No	Si	Si	No
GPR18	0	0.13	No	No	Si	No	No
GPR65	0	0.57	No	No	Si	No	No
GRP	0	0.23	No	No	Si	No	No
GUCY1A3	0.08	0.98	No	No	No	Si	No
GUCY1B3	0.23	0.98	No	No	No	Si	No
H2AFX	1	0.81	Si	No	No	No	No
HEBP1	0.85	1	No	No	Si	No	No
HGF	0.02	0.57	No	No	No	Si	No

HIST1H2BC	0.40	0.04	Si	No	No	No	No
HIST1H2BD	0.75	0.40	Si	No	No	No	No
HIST1H2BE	0.47	0.08	Si	No	No	No	No
HIST1H2BF	0.47	0.08	Si	No	No	No	No
HIST1H2BH	0.51	0.15	Si	No	No	No	No
HIST2H2BE	0.38	0.08	Si	No	No	No	No
HJURP	0.55	0	Si	No	No	No	No
HTR2B	0.21	0.66	No	No	Si	No	No
IAPP	0	0.57	No	No	Si	No	No
IGF1	0	0.94	No	No	No	Si	No
IGKV3-20	0	0.96	No	No	Si	No	No
IGKV4-1	0	0.85	No	No	Si	No	No
IL8	0.87	0.96	No	No	Si	No	No
ITPR1	0.08	0.75	No	No	No	Si	No
ITPR2	0.06	0.49	No	No	No	Si	No
KCNMA1	0.19	0.75	No	No	No	Si	No
KIF18A	0.23	0.02	Si	Si	No	No	No
KIF20A	0.85	0.09	Si	Si	No	No	No
KIF23	0.79	0.04	Si	No	No	No	No
KIF2C	0.87	0	Si	Si	No	No	No
KNTC1	0.17	0	Si	Si	No	No	No
LMNB1	0.89	0	Si	No	No	No	No
LPAR1	0.15	0.89	No	No	Si	No	No
LPAR6	0.47	1	No	No	Si	No	No
MAD1L1	0.30	0	Si	Si	No	No	No
MAD2L1	0.98	0.40	Si	Si	No	No	No
MAPK14	0.49	0.74	No	No	No	Si	No
MCM10	0.17	0	Si	Si	No	No	No
MCM2	0.74	0.02	Si	Si	No	No	No
MCM3	0.81	0.15	Si	Si	No	No	No
MCM4	0.87	0.04	Si	Si	No	No	No
MCM5	0.96	0.42	Si	Si	No	No	No
MCM6	1	0.85	Si	No	No	No	No
MCM7	0.87	0.25	Si	Si	No	No	No
MDM2	0.45	0.23	Si	No	No	No	No
MIS18A	0.45	0.02	Si	No	No	No	No
MLF1IP	1	0.51	Si	Si	No	No	No
MMP11	0	0.74	No	No	No	No	Si
MMP2	0.08	1	No	No	No	No	Si
MMRN1	0	0.11	No	No	No	Si	No
MRVI1	0	0.98	No	No	No	Si	No

NDC80	0.92	0.60	Si	Si	No	No	No
NEK2	0.98	0.15	Si	No	No	No	No
NPY1R	0.17	0.36	No	No	Si	No	No
NUF2	0.87	0.09	Si	Si	No	No	No
NUMA1	0.43	0.26	Si	No	No	No	No
NUP85	0.79	0.26	Si	Si	No	No	No
NUP98	0.64	0.32	Si	Si	No	No	No
OIP5	0.77	0	Si	No	No	No	No
OPN3	0.51	0.64	No	No	Si	No	No
OR2W3	0	0	No	No	Si	No	No
ORC4	0.74	0.57	Si	Si	No	No	No
ORC5	0.42	0.04	Si	Si	No	No	No
ORC6	0.15	0	Si	Si	No	No	No
P2RX1	0	0.11	No	No	No	Si	No
P2RY10	0	0.08	No	No	Si	No	No
P2RY12	0	0.08	No	No	Si	No	No
P2RY13	0	0.23	No	No	Si	No	No
P2RY14	0	0.23	No	No	Si	No	No
PCOLCE	0.06	0.98	No	No	No	No	Si
PDE10A	0.09	0.28	No	No	No	Si	No
PDE1A	0	0.85	No	No	No	Si	No
PDE3A	0.04	0.75	No	No	No	Si	No
PDE5A	0.13	0.92	No	No	No	Si	No
PENK	0	0.15	No	No	Si	No	No
PLCB1	0.28	0.77	No	No	Si	No	No
PLEK	0	0.47	No	No	No	Si	No
POLD2	0.19	0	Si	Si	No	No	No
POLE2	0.36	0	Si	Si	No	No	No
POLR2F	0.70	0.23	Si	Si	No	No	No
POLR2I	1	0.70	Si	Si	No	No	No
PPBP	0.07	0.15	No	No	Si	No	No
PPP2R5C	0.87	1	No	No	No	Si	No
PPP2R5E	0.19	0.74	No	No	No	Si	No
PPY	0	0.68	No	No	Si	No	No
PRKACB	0.87	1	No	No	Si	No	No
PRKAR2B	0.04	0.72	No	No	Si	No	No
PRKCB	0	0.42	No	No	No	Si	No
PRKG1	0.02	0.81	No	No	No	Si	No
PROS1	0.55	1	No	No	No	Si	No
PRSS1	0.15	0.79	No	No	No	No	Si
PSMD13	0.96	0.62	Si	Si	No	No	No

PSMD3	0.85	0.11	Si	Si	No	No	No
PSMD5	0.21	0	Si	Si	No	No	No
PSME4	0.74	0.49	Si	Si	No	No	No
PTCH1	0.02	0.87	No	No	Si	No	No
PTGER3	0	0.08	No	No	Si	No	No
PTGER4	0.78	1	No	No	Si	No	No
PTTG1	1	0.70	Si	No	No	No	No
RANGAP1	0.26	0	Si	Si	No	No	No
RFC2	0.26	0	Si	Si	No	No	No
RFC3	0.85	0.09	Si	Si	No	No	No
RFC4	0.74	0.21	Si	Si	No	No	No
RFWD2	0.79	0.40	Si	No	No	No	No
RPA3	0.94	0.51	Si	Si	No	No	No
RRM2	1	0.79	Si	No	No	No	No
RUVBL1	0.62	0	Si	No	No	No	No
S1PR1	0	0.11	No	No	Si	No	No
S1PR3	0	0.79	No	No	Si	No	No
SCG3	0	0.40	No	No	No	Si	No
SCTR	0.04	0.34	No	No	Si	No	No
SERPINA1	0.77	1	No	No	No	Si	No
SERPINE1	0.13	0.98	No	No	No	Si	No
SERPING1	0.06	1	No	No	No	Si	No
SERPINH1	0.66	1	No	No	No	No	Si
SGOL2	0.40	0.04	Si	Si	No	No	No
SKP2	0.32	0.02	Si	No	No	No	No
SLC8A1	0	0.42	No	No	No	Si	No
SPARC	0.09	1	No	No	No	Si	No
SPC25	0.60	0	Si	Si	No	No	No
SRGN	0.04	1	No	No	No	Si	No
SST	0.08	0.87	No	No	Si	No	No
STXBP3	0.79	0.98	No	No	No	Si	No
SUN2	0.57	0.36	Si	No	No	No	No
TGFB1	0.45	0.98	No	No	No	Si	No
TGFB2	0.15	0.49	No	No	No	Si	No
THBS1	0.57	1	No	No	Si	Si	No
TOP2A	1	0.60	Si	No	No	No	No
TP53	0.57	0.17	Si	No	No	No	No
TPSAB1	0	0.87	No	No	No	No	Si
TUBA4A	1	0.81	Si	Si	No	No	No
TUBG1	0.91	0.23	Si	No	No	No	No
TYMS	0.98	0.68	Si	No	No	No	No

UBE2C	0.96	0.17	Si	No	No	No	No
VEGFC	0.04	0.53	No	No	No	Si	No
VIP	0	0.09	No	No	Si	No	No
VWF	0	0.98	No	No	Si	Si	No
WNT2	0.02	0.17	No	No	Si	No	No
WNT5A	0	0.98	No	No	Si	No	No
ZWILCH	0.91	0.40	Si	Si	No	No	No
ZWINT	1	0.77	Si	Si	No	No	No

Tabla S3. Información de pase de los perfiles de expresión pertenecientes a los modelos PDX generados específicamente para este trabajo. Incluye los modelos PDX derivados de tumor primario (GSE51798) [76] y los derivados de metástasis.

Muestra	Establecimiento de los modelos PDX	Extracción del ARN e hibridación	Número de pase en JH	Número de pase en CNIO
JH029 - GSM1252898	JH	CNIO	5	1
JH051 - GSM1252903	JH	CNIO	No disponible	2
JH069 - GSM1252904	JH	CNIO	4	3
JH088 - GSM1252905	JH	CNIO	No disponible	3
JH090 - GSM1252906	JH	CNIO	3	1
JH094 - GSM1252907	JH	CNIO	5	2
JH109 - GSM1252910	JH	CNIO	3	2
JH112 - GSM1252911	JH	CNIO	4	2
JH021 - GSM1252895	JH	CNIO	9	2
JH032 - GSM1252899	JH	CNIO	4	1
JH041 - GSM1252902	JH	CNIO	No disponible	2
JH097 - GSM1252908	JH	CNIO	No disponible	2
JH102 - GSM1252909	JH	CNIO	No disponible	3
JH127 - GSM1252912	JH	CNIO	2	1
JH128 - GSM1252913	JH	CNIO	6	2
JH131 - GSM1252914	JH	CNIO	5	1
JH161 - GSM1252915	JH	CNIO	3	3
JH171 - GSM1252916	JH	CNIO	4	1
Panc007 - GSM1252917	CNIO	CNIO	–	2
Panc008 - GSM1252918	CNIO	CNIO	–	2
Panc014 - GSM1252919	CNIO	CNIO	–	4
Panc017 - GSM1252920	CNIO	CNIO	–	1
Panc023 - GSM1252921	CNIO	CNIO	–	2
Panc042	CNIO	CNIO	–	1
Panc047	CNIO	CNIO	–	2
PDXMet_Panc10	CNIO	CNIO	–	2
PDXMet_Panc11	CNIO	CNIO	–	2
PDXMet_Panc16	CNIO	CNIO	–	1
PDXMet_Panc19	CNIO	CNIO	–	1
PDXMet_Panc20	CNIO	CNIO	–	1
PDXMet_Panc25	CNIO	CNIO	–	2
PDXMet_Panc042	CNIO	CNIO	–	1
PDXMet_Panc047	CNIO	CNIO	–	2

ANEXO II – Publicaciones y comunicaciones a congresos

Artículos asociados a la tesis:

Martinez-Garcia R, Juan D, Rausell A, Muñoz M, Baños N, Menéndez C, Lopez-Casas PP, Rico D, Valencia A and Hidalgo M. **Transcriptional Dissection of Pancreatic Tumors Engrafted in Mice**. *Genome Medicine*, 2014. **6**:27.

Otros artículos:

Martinez-Garcia R, Lopez-Casas PP, Rico D, Valencia A and Hidalgo M. **Colorectal cancer classification based on gene expression is not associated with drug response.** *Nature Medicine*. En revisión.

